NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

APPLYING ENSEMBLE PREDICTION SYSTEMS TO DEPARTMENT OF DEFENSE OPERATIONS

by

Jeffrey G. Cunningham

March 2006

Thesis Advisor:	Carlyle H. Wash
Co-Advisor:	Patrick A. Harr

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** March 2006	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE: Title (Mix case letters) Applying Ensemble Prediction Systems to Department of Defense Operations			**5. FUNDING NUMBERS**
6. AUTHOR(S) Jeffrey G. Cunningham			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE**

13. ABSTRACT (maximum 200 words)

Based on recent advances, skilled objectively-determined probabilistic forecasts of some weather phenomena may be provided to operational decision-makers. Objective probabilistic forecasts that are generated from ensemble prediction systems (EPS) are attractive as a forecast methodology for Department of Defense (DoD) applications for three reasons: first, atmospheric scientists understand that the atmosphere has a limit of predictability, which means that traditional deterministic forecasts lack important uncertainty information; second, it has been demonstrated that quantifying uncertainty may improve a weather forecast user's ability to make a better decision based on their own utility function, which translates to better operational risk management (ORM) for the DoD; and finally, progress points towards a future with machine-to-machine warfare. These assertions are examined by applying probabilistic forecasts from an ensemble-based aircraft-scale turbulence forecast system to several cases and scenarios. Results clearly demonstrate the advantage of using ensemble-based probabilistic forecasts versus deterministic forecasts. Additionally, application of ensemble-based probabilistic forecast information to DoD operations is shown to be possible through its ORM programs. Specifically, air refueling scenarios are identified that demonstrate the integration of probabilistic turbulence forecast guidance into the U.S. Air Force ORM process.

14. SUBJECT TERMS Ensemble Prediction Systems, Weather Risk Management, Operational Risk Management, Air Refueling, Decision Theory, Department of Defense, Aircraft-scale Turbulence, Air Turbulence, Turbulence			**15. NUMBER OF PAGES** 149
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UL

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

APPLYING ENSEMBLE PREDICTION SYSTEMS TO DEPARTMENT OF DEFENSE OPERATIONS

Jeffrey G. Cunningham
1st Lieutenant, United States Air Force
B.S., United States Air Force Academy, 2002

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN METEOROLOGY

from the

NAVAL POSTGRADUATE SCHOOL
March 2006

Author: Jeffrey Cunningham

Approved by: Carlyle H. Wash
Thesis Advisor

Patrick A. Harr
Co-Advisor

Philip A. Durkee
Chairman, Department of Meteorology

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Based on recent advances, skilled objectively-determined probabilistic forecasts of some weather phenomena may be provided to operational decision-makers. Objective probabilistic forecasts that are generated from ensemble prediction systems (EPS) are attractive as a forecast methodology for Department of Defense (DoD) applications for three reasons: first, atmospheric scientists understand that the atmosphere has a limit of predictability, which means that traditional deterministic forecasts lack important uncertainty information; second, it has been demonstrated that quantifying uncertainty may improve a weather forecast user's ability to make a better decision based on their own utility function, which translates to better operational risk management (ORM) for the DoD; and finally, progress points towards a future with machine-to-machine warfare. These assertions are examined by applying probabilistic forecasts from an ensemble-based aircraft-scale turbulence forecast system to several cases and scenarios. Results clearly demonstrate the advantage of using ensemble-based probabilistic forecasts versus deterministic forecasts. Additionally, application of ensemble-based probabilistic forecast information to DoD operations is shown to be possible through its ORM programs. Specifically, air refueling scenarios are identified that demonstrate the integration of probabilistic turbulence forecast guidance into the U.S. Air Force ORM process.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS

ACARS	Aircraft Communication Addressing and Reporting System
ADDS	Aviation Digital Data Service
AFB	Air Force Base
AFW	Air Force Weather
AFWA	Air Force Weather Agency
AR	Air Refueling
BCs	Boundary Conditions
CAT	Clear-Air Turbulence
CDF	Cumulative Distribution Function
CVG	Convergence
CMC	Canadian Meteorological Centre
COMET	Cooperative Program for Operational Meteorology, Education and Training
CWT	Combat Weather Team
DEF	Deformation
DoD	Department of Defense
DSH	Shearing Deformation
DST	Stretching Deformation
ECMWF	European Centre for Medium-range Forecasting
EDR	Eddy dissipation Rate
EPS	Ensemble Prediction System
ESRL/GSD	Global Systems Division of the Earth System Research Laboratory
ETFS	Ensemble-based Turbulence Forecast System
FAR	False Alarm Rate
FP	Forecast Probability
FTP	File-Transfer Protocol
FSL	Forecast Systems Laboratory
GFS	Global Forecast System
GRIB	Gridded Binary
GTG	Graphical Turbulence Guidance
HIRAS	High Resolution Analysis System
HR	Hit Rate
ICs	Initial Conditions
JEFS	Joint Ensemble Prediction System
KHI	Kelvin-Helmholtz Instability
NCAR	National Center for Atmospheric Research
NCEP	National Center for Environmental Prediction
NMC	National Meteorological Center
NOAA	National Oceanic and Atmospheric Administration
NOGAPS	Navy Operational Global Atmospheric Prediction System
NWP	Numerical Weather Prediction
ORM	Operational Risk Management

OWS	Operational Weather Squadron
POD	Probability of Detection
PDF	Probability Distribution Function
PIREPS	Pilot Reports
Ri	Richardson Number
ROC	Relative Operating Characteristic
ROT	Rule of Thumb
SBCs	Surface Boundary Conditions
SOP	Standard Operating Procedure
SREF	Short-Range Ensemble Forecast
TKE	Turbulence Kinetic Energy
UCAR	University Corporation for Atmospheric Research
VWS	Vertical Wind Shear
WRM	Weather Risk Management

ACKNOWLEDGMENTS

I am deeply indebted to several people for helping me complete this thesis. First, I would like to thank my wife, family, and friends for praying for me to complete this thesis successfully and on-time. Despite being due to have our first child soon after graduation, my wife has remained a calm and positive source of motivation for me. Additionally, I am thankful for the overwhelming technical and motivational support from a variety of faculty in the Meteorology Department at the Naval Postgraduate School. My advisors, Professor Wash and Professor Harr more than tolerated my constant barrage of questions and requests for thesis meetings. Their technical expertise and guidance helped me immensely in writing my thesis. I would be remiss to fail to thank Bob Creasey and Mary Jordan for their advice on computer systems and computer programming. This thesis would not have been possible had I not received their technical instruction and help. Finally, I wish to thank Major Tony Eckel, USAF, my thesis sponsor from the Air Force Weather Agency, for his gracious expert advice.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Similar to their civilian counterparts, military meteorologists are concerned with delivering accurate and useful forecasts to their respective customers. Military forecast users have particular mission requirements, and therefore require different types of forecasts. A United States Air Force meteorologist tailors a general weather forecast to a particular weapons platform to optimize weapons payloads and to plan for possible exploitation of weather events. To provide these types of forecasts, forecasters are embedded in units that are at the "tip of the spear." A goal of integrating weather personnel into operating units is to ensure that "decision-grade environmental information for supported units" (Air Force Instruction 15-128, 2005) is incorporated into the decision-making process. Unfortunately, many of the forecast methods employed by forecasters are based on a subjective assessment of a deterministic forecast. In some cases, probabilistic forecasts may better inform decision-makers, as they convey some measure of uncertainty in the forecast. More than deterministic forecasts, probabilistic forecasts will help Air Force Weather (AFW) provide "decision-grade" information to the customer. Based on recent advances, skilled objectively-determined probabilistic forecasts of some weather phenomena may be provided to operational decision-makers.

Objective probabilistic forecasts that are generated from ensemble prediction systems (EPS) are attractive as a forecast methodology for Department of Defense (DoD) applications for three reasons: i) Atmospheric scientists understand that the atmosphere has a limit of predictability, which means that traditional deterministic forecasts lack important uncertainty information (Wilks 2006; Anthes 1986; Lewis 2005); ii), It has been demonstrated that quantifying uncertainty may improve a weather forecast user's ability to make a better decision based on their own utility function (Zhu et al. 2002), which could translate to better operational risk management for the Air Force; and iii), Progress points towards a future with machine-to-machine warfare. Machine-to-machine warfare will require data from many sources, which includes atmospheric variables. Traditional deterministic forecasts will likely be inadequate for future advanced dynamic decision-making models that likely will be inherent in advanced weapon systems. Therefore, skilled probabilistic forecasts of atmospheric phenomena that impact military

operations will be necessary. The value of the human forecaster cannot be overstated (Brooks and Doswell 1993), but a balance between automated machine forecasts and human forecasts will be needed. Also recently, the need for probabilistic forecasts has been outlined in the draft *Plan for the Joint Ensemble Forecast System* (F. Eckel, 2005, personal communication). Therefore, it is hypothesized that advancements in the area of computers, meteorology, decision theory, statistics, and weapon systems can lead to a transformation in the way military meteorologists provide forecast information for military applications.

The three main objectives of this thesis are to: (1) create an ensemble-based turbulence forecast system capable of producing forecast probability for air turbulence that impacts flight operations, (2) to demonstrate the advantages of providing forecasts based on probability of occurrence over traditional deterministic forecasts, and (3) to demonstrate the integration of probabilistic turbulence forecast information into the Air Force decision-making process.

This thesis has been organized into seven chapters: Introduction (Chapter I), Background, Methodology, Results chapters, and Conclusion. The Background chapter (Chapter II) is subdivided into three main sections that review background literature and documentation for Ensemble Prediction Systems, Weather Risk Management (WRM), and Aircraft-Scale Turbulence. These topics are further subdivided to introduce certain ideas and theories basic to understanding how to create and implement an EPS capable of producing reliable forecasts of aircraft-scale turbulence. The Methodology (Chapter III) Chapter discusses the proposed turbulence forecasting method. Three Results chapters correspond to the three main thesis objectives. The first results chapter (Chapter IV) explicitly describes the ETFS design and setup. The second results chapter (Chapter V) details the techniques and results for the second thesis objective (i.e., demonstrating the advantages of probabilistic forecast information versus deterministic forecast information). Chapter VI addresses the third thesis objective (i.e., integrating probabilistic forecast information into Air Force decision-making process). Finally, conclusions and recommendations are made in the Conclusion chapter (Chapter VII).

II. BACKGROUND

A basic understanding of meteorology (with emphasis in numerical weather prediction), decision theory, statistics, and military weapons systems is important to understand the complex problem of applying ensemble forecast products to DoD operations. Background will be explicitly given on ensemble forecasting, weather risk management, and air turbulence. It is assumed the reader will have some basic understanding of statistics and military weapon systems.

A. ENSEMBLE PREDICTION SYSTEMS

1. History and General Assumptions About Atmospheric Predictability

According to Wilks (2006), "forecasting would be… easy and meteorology boring" if the atmosphere were constant or strictly periodic, because describing it mathematically would be easy. The atmosphere is neither constant nor strictly periodic. Based on the literature, a consensus is developing when considering atmospheric uncertainty. First, "dynamical chaos," as defined by Lorenz (1963), is inherent to the atmospheric system (i.e. atmospheric non-linear, dynamic equations are highly sensitive to initial conditions) and so even if the models had perfect physics and dynamics, there would still be uncertainty in the forecasts (Wilks 2006). Second, models do not have perfect dynamics and physics and therefore there exists some model error that increases atmospheric forecast uncertainty (Wilks 2006; Anthes 1986). Wilks (2006) asserts, "deterministic forecasts of future atmospheric behavior will always be uncertain, and probabilistic methods will always be needed to adequately describe that behavior."

The most accurate approach to providing probabilistic forecast data would be to use stochastic dynamic prediction. One would apply the deterministic equations, those equations that define the laws governing the atmosphere, to the initial condition probability distributions describing the uncertainty in the initial state of the atmosphere. The process would yield forecasts that are probability distributions of the future state of the atmosphere (Wilks 2006). In theory, the stochastic-dynamic prediction approach is appealing. Practically, the current sets of equations used to define the atmospheric laws

are inadequate. Additionally, representing the millions of dimensions for the phase state of an atmospheric system would require extreme amounts of computing power.

A more practical approach is to approximate the pure stochastic approach with an EPS. Initially stochastic-dynamic prediction was conducted using Monte Carlo methods suggested by Esptein and Leith (Lewis 2005). Monte Carlo methods assume a known randomly sampled probability density function (PDF) (Lewis 2005). However, operationally derived perturbations are produced through singular vector or breeding vectors, which are not random (Lewis 2005). The need for using ensemble methods was generated by the inherent sampling problem with Monte Carlo methods (Lewis 2005). Lewis (2005) notes, "…it remains to be determined the most appropriate way to perturb the models..." Much work is being done to improve how EPS systems perturb ensemble members.

Kalnay (2003) points out that ensembles primarily differ in how they generate initial perturbations. She classified the methods as, "those that have random initial perturbations and those where the perturbations depend on the dynamics of the underlying flow." Monte Carlo methods are in the first class and bred and singular vectors are in the second class. The methods in the second class rely on the "errors of the day." She notes that multi-model (models from different operational centers) and multi-data assimilation techniques are promising methods as well.

2. Basic EPS Design and Variations

An EPS is designed to account for two areas of uncertainty: initial condition uncertainty and model uncertainty (error). Essentially, an EPS may have: 1) ensemble members that vary initial conditions/boundary conditions (ICs/BCs); 2) ensemble members that vary by model-related errors (i.e., different numerical schemes, parameterization, etc.); or 3) members that combine both methods (Toth and Vannitsem 2005).

Initially, ensemble prediction systems produced forecasts by only varying ICs. This method of producing ensembles became widely accepted at many national weather centers across the world and still remains the primary method for at least two global ensembles: the National Centers for Environmental Protection (NCEP) Ensemble and the

European Centre for Medium-range Forecasting (ECMWF) Ensemble. The NCEP ensemble and the ECMWF ensembles differ in how they perturb initial conditions. The NCEP ensemble uses a bred vector method and the ECMWF ensemble uses a singular vector approach (Kalnay 2003). Perturbing only the ICs is valid if model-related errors do not dominate the final error fields (Toth et al., 1997). The Canadian Meteorological Centre (CMC) EPS takes into account the IC and model-related error. In addition to having perturbed observations, some of the 16 members of the ensemble have different physics packages. The CMC EPS also has perturbed boundary conditions "such as sea surface temperature, albedo and roughness length" [CMC Website Available online at: http://weatheroffice.ec.gc.ca/ensemble/index_e.html (Current as of June 17, 2005)].

Eckel and Mass (2005) have also delineated methods and terminology associated with ensemble prediction systems. They use "multi-analysis" to describe an ensemble system with varied ICs (and lateral boundary conditions for mesoscale ensembles). Varied surface boundary conditions (SBCs) could also be lumped under the multi-analysis classification, but are often model-dependent. They note that two methods have emerged to account for model error. The first method, called model diversity, consists of two techniques. The first technique, called multi-model, utilizes completely different models to cause ensemble members to have different model attractors. The second technique, called varied-model, uses the same model, but with "…varied combinations of model physics and/or perturbed parameterization." The second method described by Eckel and Mass (2005) is called stochastic physics, in which random errors are added to the evolving solution during the model integration. A variety of strategies and methods have been developed to account for IC and model uncertainty. The development of ensemble prediction systems will continue to improve as better methods for stochastic-dynamic forecasting are found and integrated into numerical weather prediction (NWP).

3. Types of Ensemble Products/Graphics

The methods with which stochastic data are conveyed to a forecaster or forecast user will affect how well the ensemble forecast data are incorporated into the decision-making process. A good review of the known methods for displaying ensemble forecast data can be found in the University Corporation for Atmospheric Research (UCAR) website under Cooperative Program for Operational Meteorology, Education and

Training (COMET) program training module titled: “Ensemble Forecasting Explained” (UCAR 2005c). The document describes three basic types of ensemble forecast products: mean with spread, spaghetti chart, and probabilistic product types. The three product types show the “middleness and spread”, “the probability distribution of ensemble forecasts,” and “the probability of exceeding specific thresholds,” respectively (UCAR 2005c).

a. Mean and Spread

This type of product is the most compact and easiest to interpret of the three ensemble product types. The mean is the average of all ensemble members and the spread is the standard deviation of the ensemble members, which assumes a Gaussian distribution. This type of product can be used to quickly ascertain the uncertainty of the forecast. If the product indicates a high spread (high standard deviation) then there is more uncertainty in the forecast (UCAR 2005c). The mean and spread product is not good for guiding a forecaster’s conceptual model, because the mean of the ensemble members is much smoother than a single ensemble member. The smoothing effect could “smooth-out” important atmospheric features needed for proper meteorological forecasts.

b. Spaghetti

The spaghetti type of ensemble forecast product displays all ensemble members on one product, or the distribution of the members. This product can be beneficial for a quick look at where the members are in agreement or disagreement, but can become very confusing with many ensemble members. Only an incomplete assessment of the probability distribution can be determined from the spaghetti chart (UCAR 2005c).

c. Probabilistic

UCAR (2005c) illustrated two types of probabilistic ensemble forecast products, “the most likely event” and “the probability of exceedance” products/graphics. For example, the “most likely event” product can be used to distinguish precipitation type or turbulence intensity. The ensemble data would need to be post-processed through an algorithm specifically designed for a particular weather element forecast, such as precipitation type or turbulence intensity. The “most likely event” product could hide other events with nearly the same likelihood of occurrence. The “probability of

exceedance" product is good for determining the likelihood of exceeding warning thresholds. Both types of probability products do not convey the full probability distribution and may hide other possible solutions that may impact the forecast user (UCAR 2005c).

B. WEATHER RISK MANAGEMENT

A primary motivation for a probabilistic forecast approach is that it can couple uncertainty data with the risk tolerance of weather forecast users. As noted by Zhu et al. (2002), "Quantifying forecast uncertainty with an ensemble approach can improve the user's bottom line." Weather is an important factor in decision-making for many organizations and individuals. Often these organizations and individuals, hereafter called forecast users, would prefer to have advanced warning of weather phenomena (i.e., a weather forecast). The methods in which forecast users choose to integrate these data into their decision-making process vary and generally depend on the forecast user's sensitivity to weather elements.

Forecast users choose to react to a weather forecast based on two factors. First, the forecast user examines their own sensitivity to the weather phenomena with reference to a utility function. This may be subjectively or objectively determined, plus it may be in less quantifiable terms, such as life and safety. Second, the forecast user subjectively determines whether or not to trust the forecast based on their own 'fuzzy' interpretation of the certainty expressed by the forecast or forecaster. For an Air Force meteorologist, that means a pilot receiving an aviation forecast will look the forecaster in the eyes and say "Are you really certain about the forecast?" If the forecaster blinks, then the pilot knows the forecaster does not have high confidence in his or her forecast. Wilks (2006) explains that before a forecaster should report a "subjective degree of uncertainty as part of a forecast," a forecaster needs internally to develop a subjective probability distribution of their uncertainty. A reasonable estimation of objective forecast certainty can be given with a well-designed EPS. This uncertainty information should be incorporated into the decision-making process in terms of forecast probability of occurrence.

1. Cost/Loss Analysis and Economic Value

For weather forecasts to be effective, they must provide some economic value, save lives, enhance quality of life, or provide some benefit to the forecast user based on their utility functions. The study of weather forecasts with respect to economic value has been the subject of economists, meteorologists, and decision theorists for some time. Despite the obvious connection between weather and economics, Palmer (2002) notes that there is a difference in the way that weather forecasts are assessed by model developers and forecast customers, "root-mean square error of 500 hPa height on the one hand; pounds, euros, or dollars saved on the other."

Zhu et al. (2002) demonstrate the economic value of ensemble forecasts versus the traditional deterministic forecast. They believe an essential question to presenting any new method is does the new method provide higher quality guidance than the existing method? With ensembles, they demonstrated that the answer is yes, particularly at longer forecast intervals. The authors performed a detailed analysis of economic value versus cost/lost ratio (C-L ratio) and relative operating characteristics versus lead time for 500mb heights. They noted that beyond a 4-day lead time, the lower horizontal resolution T62 model ensemble outperformed the higher horizontal resolution control model. Both model formulations had similar computational costs. They concluded that for most users the ensemble offers more economic value than a single deterministic control forecast.

Both Zhu et al. (2002) and Palmer (2002) suggest that if a user's exposure to risk can be quantified and related to their risk tolerance, then better (more economical) decisions can be made. This is demonstrated by the following example from Palmer (2002). If a traditional deterministic 5-day forecast indicated benign conditions that does not mean there is not a chance of severe weather. If that deterministic forecast was all that was provided to make a decision on whether or not to tow an oil-rig to a site, a rational decision would be to proceed. If the EPS predicted a 20% chance of severe weather, the decision makers would then consider the cost of keeping the ship in port (waiting for favorable weather conditions) to the loss associated with an oil rig in-tow during a storm (i.e., cost-loss ratio).

Zhu et al. (2002) and others have demonstrated that ensembles are valuable tools for decision-making. The intrinsic value of ensemble prediction systems is the ability to ascertain some level of uncertainty. The above business logic relies on the subtle assumption that the EPS operates as a pure stochastic-dynamic model and that its forecasts encompass truth. Obviously, this is not true because models do have errors in their numerics, dynamics, physics, etc. and we are unable to sample the true PDF that represents the initial conditions. Additionally, the added benefit of ensembles is not immediately clear, unless one knows the forecast user's C-L ratio. The C-L ratio is often difficult to determine for many DoD forecast users. Instead, forecast users make decision based on a more qualitative assessment through operational risk management (ORM).

2. Operational Risk Management

An economic argument for the use of probabilistic forecast information may not be sufficient for military planners. Certainly, military planners do want to save money, but sometimes saving lives or mission success is more important than the "bottom line." Fortunately, the advantage of using a C-L analysis does not need to be confined to monetary values. One could describe C and L in terms of benefits and risks. For example, military planners balance risk with mission priority. In order for probabilistic forecast information to be accepted by DoD decision-makers, it must be integrated into the ORM decision-making process.

U.S. Air Force operational risk management guidelines and tools are defined in Air Force Pamphlet 90-902 (AFPAM 90-902). The introduction states that, "All US Air Force missions and our daily routines involve risk." In response, U.S. Air Force sub-communities further develop specific ORM worksheets and tools to aide in making operational decisions. "The USAF aim is to increase mission success while reducing the risk to personnel and resources to the lowest practical level in both on- and off-duty environments" (AFPAM 90-902). ORM is a way of life in the U.S. Air Force.

The military is very similar to its civilian business counterparts in that its goals are to preserve its people and assets. However, the two communities differ in their end goal. Ultimately, the military seeks to maximize its combat capability (AFPAM 90-902) and businesses seek to maximize their profits. The risk management goals of the U.S. Air Force can be found in Figure 1. Other DoD components follow similar ORM

guidelines and goals. In addition to the goals of ORM, there are four main principles to include: 1) accept no unnecessary risk, 2) make decisions at the appropriate level, 3) accept risk when benefits outweigh the cost, 4) integrate ORM into Air Force doctrine at all levels (AFPAM 90-902).

Figure 1. U.S. Air Force Risk Management Goals (from AFPAM 90-902).

Every mission and function of the U.S. Air Force has its own unique set of risks. Historically, Air Force Weather has only produced deterministic forecasts, which do not convey forecast certainty. This important missing information is potentially useful in light of the U.S. Air Force's guiding ORM goals and principles. Probabilistic forecasts provide the missing uncertainty information. Probabilistic forecast give the forecast user another tool to mitigate risk to the level necessary for their unique set of characteristics. Ensemble-based probabilistic forecasts can be effectively applied to a variety of DoD operations by integrating the probabilistic forecasts into the ORM process.

C. AIRCRAFT-SCALE TURBULENCE

1. The Phenomena

Atmospheric turbulence is a critical micro-to-mesoscale weather element that affects aviation-related activities in all stages of flight from takeoff to landing. Scientists and researchers have made considerable efforts to better understand, observe, and forecast turbulence. In this thesis, it is hypothesized that ensemble forecasting can improve turbulence forecasting and minimize the effects of turbulence on aviation.

However, significant integration of ensemble-based forecast probability of occurrence of turbulence into the aviation decision-making process will be necessary for substantial impact. A sufficient understanding of turbulence and how it is observed is helpful, if not critical, for any turbulence forecasting method.

Air turbulence impacts both military and civilian aviation, sometimes even causing fatalities (Ellrod and Knapp 1992). Air turbulence that affects military aviation will be addressed in this thesis. Although, similar effects may be felt in civilian aviation, the effects of air turbulence on civilian aviation will not be specifically addressed in this thesis. The effects of air turbulence on military aviation are mission-dependent. For example, air turbulence can cause aircraft that are conducting air refueling missions to reroute, change altitude, or loiter to find military operating areas that are unaffected by air turbulence. Air turbulence affects other mission types (e.g., cross-ocean transports, shuttle, etc.), as well. Finally, air turbulence adds another element of risk into an operational risk management formula. Rerouting and loitering due to air turbulence and flying through air turbulence can increase fuel expenses (Ellrod and Knapp 1992) and increases mission times at the expense of the DoD budget and U.S. national security. Improvements in air turbulence observing and forecasting are needed to positively affect the “bottom line” of the DoD budget and positively affect national security and operational risk management.

MacCready (1964) defines turbulence as “…motions at various intensities and scales in three dimensions…” and that “all the statistical properties of atmospheric turbulence can be related to one parameter, ε, a dissipation rate of turbulent energy.” He further explains that, fortunately, the inertial sub-range of ε includes the gusts that affect aircraft (aircraft fatigue problems and the human “feel” of turbulence). Ideally, NWP models would directly forecast turbulence. Unfortunately, the horizontal and vertical resolution required to do this for aircraft-scale turbulence remains too high for current NWP models. Instead, diagnostics have been developed to calculate turbulence intensity or likelihood of turbulence. Essentially, the diagnostics are algorithms or indices that parameterize turbulence for an entire grid space. A problem with these diagnostics is that they do not account for all types of turbulence. A diagnostic that is designed for clear-air turbulence may not work for mountain-wave turbulence or convection-related turbulence.

To appropriately forecast turbulence, all methods of turbulence generation should be taken into account (frontogenesis, convection, orography, etc.). The issue of appropriately forecasting turbulence will be discussed later.

2. Observing Air Turbulence

Observing air turbulence that impacts flight operations is challenging. Air turbulence is of small enough scale to require an observing system with very high horizontal and vertical resolution (probably less than one kilometer horizontal resolution) for the phenomena to be observed accurately. Until recently, most turbulence data were provided via subjective verbal pilot reports (PIREPS) at the discretion of the aircrew, which has made PIREPS a difficult tool to use for the verification of turbulence (Cornman et al 1995; Schwartz 1996; Tebaldi et al. 2002). Therefore, more objective automated turbulence measurements from aircraft are a welcome observing tool for atmospheric researchers and operational meteorologists alike. As reported in the National Center for Atmospheric Research (NCAR) Research Application Programs (RAP) 2004 Annual Report, automated turbulence measurements from aircraft, in combination with Doppler ground-based radar, are being developed as a method for clear-air turbulence observing and nowcasting (UCAR 2005b).

a. Automated Aircraft Turbulence Measurements

The Global Systems Division of the Earth System Research Laboratory (ESRL/GSD), formerly Forecast Systems Laboratory (FSL), has taken a lead role in providing automated meteorological reports from commercial aircraft to atmospheric researchers and to government operational forecasters. Recently, ESRL/GSD added automated turbulence data to the other weather data on their unofficial (not operational) website http://acweb.fsl.noaa.gov/ (ESRL/GSD 2005a; ESRL/GSD 2005b).

Automated weather reports from commercial aircraft have been assimilated into NWP models for over a decade. More recently, these data have been provided to forecasters and other users through ESRL/GSD’s website, although the data were proprietary. ESRL/GSD’s users are bound by an agreement not to release the data real-time to non-participating airlines. The data can only be used by government forecasters, such as National Oceanic and Atmospheric Administration (NOAA), and

cannot be released to airlines that do not participate in the Aircraft Communication Addressing and Reporting System (ACARS, ESRL/GSD 2005a).

The automated turbulence measurements by aircraft are estimates of a form of ε, which is MacCready's proposed universal turbulence standardization technique. It is quantitatively based on atmospheric turbulence, as opposed to the qualitative and aircraft-dependent turbulence a pilot may "feel" (MacCready 1964). In his paper, MacCready defines eddy dissipation rate (EDR) as "the rate at which the turbulence energy is converted into heat for steady turbulence." He stated that an eddy dissipation rate can be measured independently of aircraft type or speed. Eddy dissipation rate can be measured by detecting "…the small longitudinal (or lateral) velocity turbulent fluctuation…" (MacCready 1964).

According to the 2003 NCAR RAP Annual Report, the current EDR algorithm implemented on United Airlines aircraft estimates EDR turbulence intensity indirectly through vertical acceleration measurements in combination with a model of aircraft response to turbulence (UCAR 2005a). A future method will be implemented that will estimate EDR directly by estimating the vertical component of the wind vector (UCAR, 2005a). NCAR continues to conduct research sensors to better measure EDR.

It may be possible for the eddy dissipation rate to be directly ingested into NWP models, which suggests a future possibility of forecasting EDR directly (AMS 2003). If a pilot is provided EDR data directly, he/she may be able to relate that information to an aircraft-dependent chart (particular to their aircraft flight characteristics, as well) and make a determination on how to continue their flight.

At the time of this research, the use of automated turbulence observations for verification of turbulence diagnostics appears to be limited. Tebaldi et al. (2002) used vertical accelerometer data "avars" only when the reports were for null. Null reports were determined to be unambiguous. They noted that reports of actual turbulence could have been pilot induced and not a result of actual turbulence. Although automated turbulence observations were not used in this research, they would be a valuable source of verification data if used in future work.

b. PIREPS

Schwartz (1996) provided a critical review of the use of PIREPS quantitatively in developing aviation weather guidance products. In his review, he noted that the difficulties and inadequacies of using PIREPS alone for verification of aviation forecasting techniques are numerous. He reviewed several articles detailing different aviation forecasting techniques. One researcher noted that his particular forecasting method gives the probability of the reporting of clear air turbulence (CAT), not the probability of CAT. Schwartz (1996) noted that this problem doesn't render the techniques useless, just that PIREPS are not ideal for verification. He further commented that, "Familiar classic statistical measures of performance for forecasting algorithms, such as the false alarm ratio, probability of detection, and threat scores, have limited applicability when poorly observed data are used." There are other problems associated with the "nonconformity to the regulations" and "non-standardization" of reporting by pilots (Schwartz 1996). Despite the disadvantages of using PIREPS for the verification of turbulence forecasting techniques, few other better options are currently available to researchers and operational meteorologists. Automated turbulence observations, however, from aircraft will enhance the available PIREP database.

3. Forecasting Air Turbulence

No single universal method is employed by operational aviation forecast centers and DoD meteorologists for automated or manual forecasts of aircraft-scale turbulence. Most centers still provide some form of traditional human turbulence forecasts. This subsection on forecasting air turbulence will highlight several of the ways current operational centers and meteorologists forecast aircraft-scale turbulence, as well as mention a few general turbulence indices that are easily employed during post-processing to create automated turbulence forecasts.

a. Turbulence Diagnostics

Over the last 50 years, many methods have been proposed to forecast air turbulence. Only some of the methods that deal with clear-air turbulence (CAT) will be discussed here. Tebaldi et al. (2002) reviewed many of the turbulence diagnostics developed over the years, and some of the known turbulence diagnostic methods are listed here:

- Vertical wind shear
- Horizontal wind shear
- Richardson number
- Turbulence Kinetic Energy (TKE)
- Colson-Panofsky index
- Ellrod indices
- Enlich empirical wind index
- Brown's index
- Reap MOSS predictors
- Dutton's empirical index.

A complete list (in detail, with equations) can be found in Tebaldi et al. (2002). Each method of forecasting turbulence varies in approach. Some methods consider vertical wind shear, horizontal wind shear, stability, and/or vorticity. They compared the performance of the different indices with the same dataset and found that some of the indices consistently performed poorly, and therefore suggested disregarding those indices. Although, the TKE preformed best, they concluded that "indices considered in isolation are not very informative, and that a multidimensional approach performs better in predicting CAT" (Tebaldi et al 2002). In addition to examining the index approaches individually, they applied different multivariate techniques to predict turbulence and found that the multivariate model "borrows strength across the different predictors" (Tebaldi et al. 2002). A pseudo-multivariate approach in conjunction with an EPS will be proposed later as a possible method for forecasting the probability of occurrence of turbulence.

A diagnostic that uses spatial structure functions of model variables (such as velocity fields and temperature) to estimate small-scale turbulence has been proposed by Frehlich and Sharman (2004). This method seems promising because of the ability to use model output to accurately represent small-scale turbulence at sub-model grid scales. The method also promises to positively contribute to building a climatology of turbulence (Frehlich and Sharman 2004).

b. Aviation Weather Center (AWC)

An automated product produced by RAP, which is available as a nowcast and forecast product through the AWC and distributed through the Aviation Digital Data Service (ADDS) website [Available online at: http://adds.aviationweather.noaa.gov/turbulence/], is called Graphical Turbulence Guidance (GTG). The GTG product appropriately weights the forecast with PIREPS. EDR data reported automatically from aircraft are being introduced into the process (UCAR, 2005b). RAP NCAR scientists are encouraged with the new diagnostic developed by Frehlich and Sharman (2004), since it can produce model-derived EDR fields (UCAR 2005b).

c Air Force Weather (AFW)

Within AFW, several methods exist by which turbulence forecasts are generated and disseminated. AFW operates with a forecast-funnel approach using three levels (ideally, higher-level forecasts guiding lower-level forecasts): the strategic level, the operational level, and the tactical level. Strategic-level forecasts are created at the Air Force Weather Agency (AFWA) at Offutt Air Force Base (AFB), Nebraska (Air Force Instruction 15-128, 2005). Operational-level forecasts are created at operational weather squadrons (OWSs) throughout the world (i.e., regional hubs). Tactical-level forecasts are issued at the base level by combat weather teams (CWTs). Each level issues their own forecasts based on their area of concern and their operational level. The strategic-level forecasts normally are global or hemispheric in nature and are typically automated model guidance issued from the modeling branch at AFWA. OWSs issue forecasts for their specific region (e.g., traditional human graphic charts). Finally, CWTs issue forecast to the war fighter (e.g., pilot) in the form of a written one page text forecast describing where they will experience turbulence on their mission. CWTs also have various ways of giving weather briefings tailored to their specific customer. Some of the regional charts produced by the OWSs are often included in the pilot weather briefing.

Turbulence forecast products are issued in a similar manner from each of the three levels of operation. At the strategic level, AFWA produces automated upper-level turbulence forecast guidance (note Figure 2) based on post-processed MM5 output using the second version of Ellrod's objective CAT turbulence index, as defined in Ellrod and Knapp (1992) (G. Brooks, 2005, personal communication). Additionally, AFWA

produces low-level turbulence forecast guidance based on post-processed MM5 output using an AFWA modified version of the Panofsky index (G. Brooks, 2005, personal communication). Operational-level forecasts of turbulence are generally traditional human forecasts in regional chart form. The forecasters at the OWS who produce the turbulence forecast are trained to use model guidance, rules-of-thumb (ROTs) from the Air Force Weather Agency Technical Note 98-002, and local standard operating procedures (SOPs). The ROTs are generally based on studies from the 1960s that were based on synoptic factors. Therefore, OWS turbulence forecasts are subjective in nature (note Figure 3). Finally, CWTs produce tailored text turbulence forecasts to pilots (Flight Weather Briefing – note Figure 4) and may attach a copy of the regional turbulence forecast chart produced by the OWS. CWTs provide subjective human forecasts that are tailored to the mission requirements of the customer.

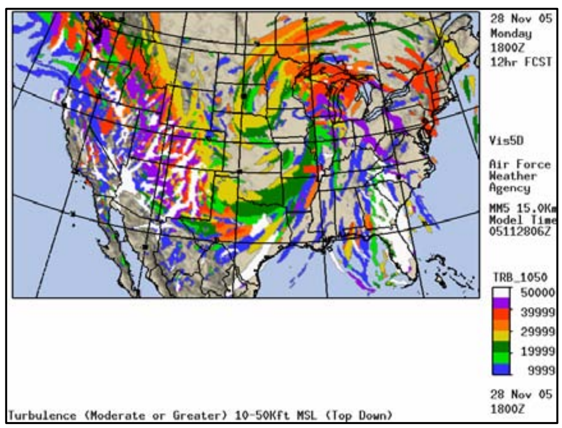


Figure 2. Example AFWA Model-Derived Turbulence Forecast (from JAAWIN 2005).

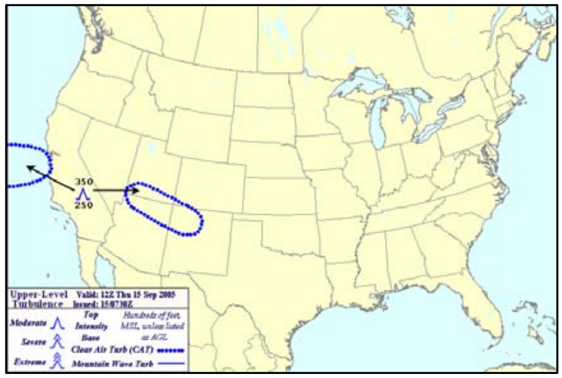


Figure 3. Example OWS Human Turbulence Chart (from Barksdale OWS 2005).

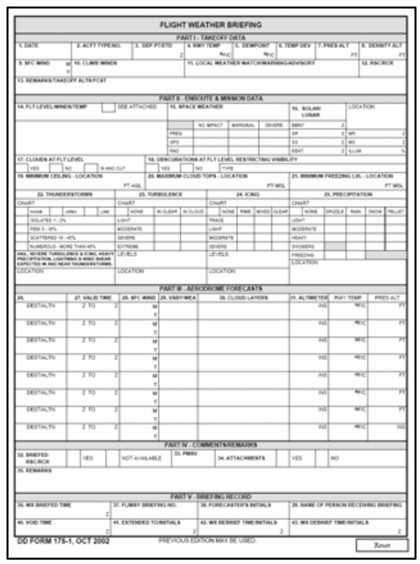


Figure 4. Example Flight Weather Briefing (from DD175-1 2005).

d. Short-Range Ensemble Forecasting System (SREF) Aviation Project

Under the sponsorship of the Federal Aviation Administration (FAA) the NCEP began the SREF Aviation Project to provide mesoscale probabilistic forecast information to the aviation world (Zhou et al. 2004). Forecast probability of turbulence occurrence is one of the many experimental forecast probability products created by the SREF Aviation Project. These products employ the simple Ellrod turbulence diagnostic index (Zhou et al. 2004), which is easy to use in conjunction with probability of

exceedance type ensemble forecast products. Zhou et al. (2004) note that verification of aviation-related parameters has yet to occur because of the difficulties applying their current deterministic verification tools to probabilistic forecasts.

THIS PAGE INTENTIONALLY LEFT BLANK

III. METHODOLOGY

Using the previous background research as a foundation, it is hypothesized that a well-designed ETFS based on existing EPSs [e.g., NCEP's Global Forecast System (GFS) ensemble, U.S. Navy's Operational Global Atmospheric Prediction System (NOGAPS) ensemble, etc.] would improve AFW's ability to forecast air turbulence, maximize cost effectiveness, and positively affect U.S. national security by increasing mission effectiveness.

A well-designed EPS captures analysis uncertainty and model uncertainty. For an ETFS to capture analysis uncertainty, it will obviously need to be multi-analysis. To capture model uncertainty, it will need to be varied-model (varied turbulence diagnostics) and possibly multi-model (varied core model). Tebaldi et al, (2002) indicated that some turbulence diagnostic methods yield better results than others, so calibrated (weighted) ensemble members would be required for a skilled ensemble. Weighting factors could be based on how well the particular members have performed over some period of time. Different ensemble members could be based on different turbulence diagnostics. However, Tebaldi et al. (2002) mention that there could be a problem of calibrating the different diagnostic methods because of the difficulty of verifying air turbulence. A possible solution may be to use the EDR fields defined by spatial structure functions, as defined by Frehlich and Sharman (2004) to estimate the climatology and for predicting small-scale turbulence below the model grid scale. An effective ETFS would also forecast more than one type of turbulence (i.e., mountain wave turbulence, convective turbulence, etc.).

To benefit from the power of a well-designed ETFS, methods for integrating ETFS probabilistic output (note Figure 5) into the AF decision-making process need to be created to convey uncertainty information to the forecast user in an effective and beneficial manner. At some point, the following questions need to be answered: "When in the decision-making process is probabilistic information most needed? When is it most effective? How should it be conveyed? If the forecast users have an automated decision-making process, how can we incorporate stochastic forecasts into the process?"

Interestingly, Dutton (1980) had already asserted that forecasts of CAT must be stated in terms of probability to convey “maximum possible information to the user.”

To answer some of these questions, the author created a rudimentary ETFS based on GFS ensemble output and Ellrod’s Turbulence Index to apply to real world AF scenarios. Figure 5 is an example stochastic turbulence forecast product produced by the ETFS created for this thesis. Figure 5 represents the forecast probability of moderate to severe turbulence for the layer 30,000 ft to 39,000 feet. Warm colors represent a high probability of moderate to severe turbulence for the given layer represented in the figure. Cool colors represent lower probabilities of moderate to severe turbulence.

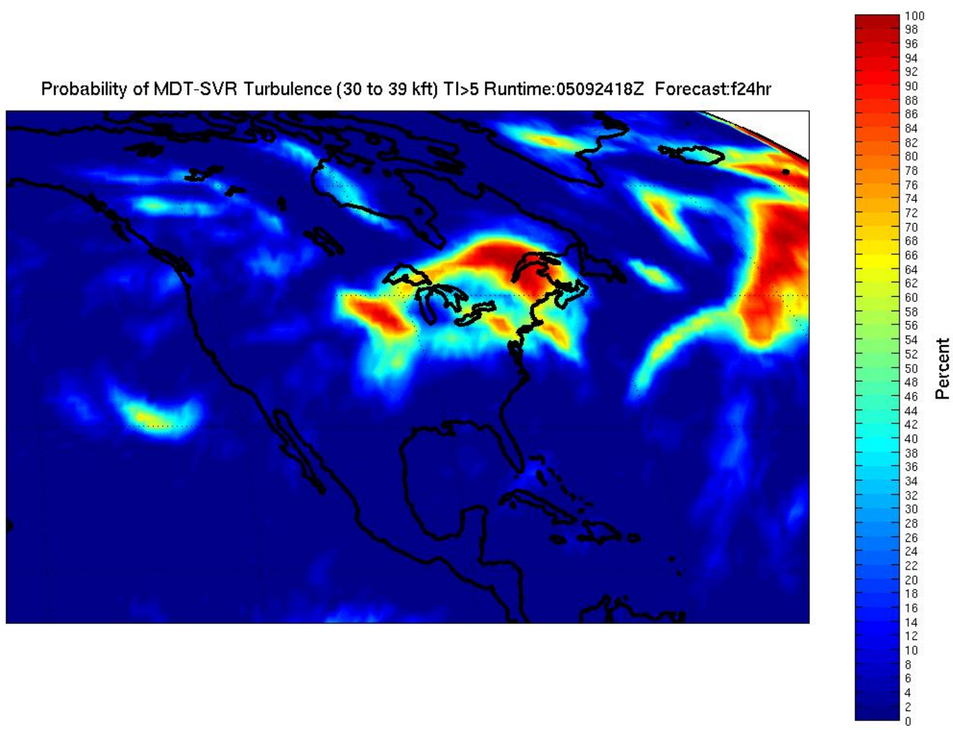


Figure 5. Example Turbulence Probability of Occurrence from ETFS.

The following three chapters report the techniques used and the results obtained for the three main objectives of this thesis. Recall, the three main objectives of this thesis are to: (1) create an ensemble-based turbulence forecast system capable of producing forecast probability for air turbulence that impacts flight operations, (2) to demonstrate

the advantages of providing forecasts based on probability of occurrence over traditional deterministic forecasts, and (3) to demonstrate the integration of probabilistic turbulence forecast information into the Air Force decision-making process.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. AN ENSEMBLE-BASED TURBULENCE FORECAST SYSTEM (ETFS)

Many considerations must be weighed when designing an EPS. Assumptions about initial condition uncertainty and dynamic or physics uncertainty are primary 'theoretical' drivers behind many EPS design choices. However, there are many practical and technical limitations that must be taken in to account when designing an EPS system. For example, a perfect ensemble would need an infinite number of ensemble members, but practically only a discrete number of ensemble members can be created due to limited computer resources. Other conflicts between theoretical motivations and practical limitations will be described later in this section. All programming for this ETFS and thesis were conducted in Matlab.

In this study, the ETFS will be used as a tool to produce forecast probability of occurrence of moderate to severe aircraft scale turbulence. The forecast probability created by the ETFS will be used to demonstrate the advantages of using forecast probability over deterministic forecasts in decision-making.

A. GLOBAL FORECAST SYSTEM (GFS) ENSEMBLE

The GFS ensemble model output available from NCEP's file transfer protocol (FTP) servers, which is available in gridded binary (GRIB) format, serves as the basis for producing ensemble-based aircraft-scale turbulence forecasts. The goal of a good EPS is to produce a reasonable random sample of the real distribution. NCEP uses the breeding method to create individual ensemble members.

B. TURBULENCE DIAGNOSTIC

The turbulence diagnostic is based on the Ellrod turbulence index (Ellrod and Knapp 1992). This method was chosen for its ease of implementation and because NCEP and AFWA have used this diagnostic with model output with some success over the last decade or so. Ellrod and Knapp (1992) describe their method as "an objective clear-air turbulence forecasting technique." However, as mentioned in the background section, using a single turbulence diagnostic may not be as informative as using a multivariate

approach. Thus, it should be understood that using the Ellrod turbulence diagnostic alone forces the ETFS to be rudimentary, at best.

The Ellrod index was developed with the backdrop of turbulence studies conducted from the 1950s to the 1980s. Many of the studies in the 1950s and 1960s focused on CAT and found that the principal mechanism responsible for CAT was Kelvin-Helmholtz instability (KHI) (Ellrod and Knapp, 1992). They noted that “KHI occurs when vertical wind shear within a stable layer exceeds a critical value.” In addition, they commented that the Richardson number (Ri) has also been used, but fails to be operationally useful. Forecast offices first attempted to forecast turbulence by determining synoptic and mesoscale conditions favorable for turbulence (Ellrod and Knapp, 1992). Early on, it was noted that aircraft scale turbulence is on too small a scale to be resolvable by numerical weather prediction models. This continues to be a problem today. Synoptic and mesoscale flow patterns were empirically related to patterns of the occurrence of CAT in order to compensate for the inability to directly forecast turbulence.

The physical basis for the Ellrod index is based on a Petterssen equation for frontogenetic intensity. Ellrod and Knapp (1992) note that frontogenesis increases vertical wind shear, which increases the likelihood of CAT occurrence. However, Ellrod’s turbulence index has a strong dependence on the product of both deformation (DEF) and vertical wind shear (VWS). Both DEF and VWS can be simply calculated by the using u and v wind-component forecasts. Ellrod and Knapp (1992) found that the product of VWS and DEF resulted in a higher correlation to the occurrence of CAT than DEF alone. Ellrod and Knapp (1992) define VWS as a rapid change in wind speed and/or direction with height. The following equation is used to define VWS:

$$VWS = \frac{(\Delta u^2 + \Delta v^2)^{1/2}}{\Delta z} \tag{1}$$

Δz is the layer thickness. Refer to the pressure column in Table 4 for upper and lower bounds of each layer by pressure. The calculated geopotential height for the given pressure and grid point was used for actual calculations. Δu and Δv are the difference between wind speeds for u and v, respectively. Both Δu and Δv were calculated at a

middle layer between the upper and lower pressure level bounds. For example, for the 200 to 300 mb layer (Layer 1 from Table 4), Δu and Δv were calculated at the 250 mb level.

Ellrod and Knapp (1992) define deformation as “a property of a fluid that transforms a circular-shaped area of fluid to an elliptical shape.” DEF is defined as:

$$DEF = (DST^2 + DSH^2)^{1/2} \tag{2}$$

where DST is the stretching deformation and DSH is the shearing deformation. They are defined as:

$$DST = \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \tag{3}$$

$$DSH = \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \tag{4}$$

Ellrod and Knapp (1992) define two versions of their index, TI1 and TI2. In equation form, they are:

$$TI1 = VWS\ X\ DEF \tag{5}$$

$$TI2 = VWS\ X\ [DEF + CVG] \tag{6}$$

They state that convergence (CVG), “a compaction of a fluid caused by the confluence of streamlines and/or deceleration of air parcels,” contributes to frontogenesis. CVG is negative divergence and is defined as

$$CVG = -\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right). \tag{7}$$

In addition, they noted that strong subsidence causes turbulence in some cases. TI2 was chosen for this research, since AFWA currently uses this index for its automated turbulence forecasts based on the MM5 (G. Brooks, 2005, personal communication). The above equations were applied by using a finite differencing approach during post-processing on grids from GFS ensemble members.

Unfortunately, the Ellrod turbulence indices (and turbulence indices, in general) are difficult to verify. As noted earlier, PIREPS are not adequate for verifying turbulence. Automated turbulence observations help enhance the existing PIREP database, but may not be sufficient for verifying turbulence diagnostics. Before use as a skillful turbulence forecasting method, the Ellrod indices require calibration. Ellrod and Knapp (1992) note that their indices are model-dependent. Since the Ellrod indices are applied with model output on different grid resolutions, results may vary between models with different grid resolution. To the author's knowledge there haven't been any studies done to demonstrate the usability of the index at different resolutions. One must realize that using this index on even high-resolution grids is still only parameterizing aircraft scale turbulence, not directly calculating the phenomena.

Despite the inadequacies of the current turbulence observing network, some level of validity may be ascertained from the data. Ellrod and Knapp (1992) subjectively determined intensity thresholds (e.g., light, moderate, severe, etc.).

C. ELLROD INDEX THRESHOLD CALIBRATION

1. Overview

Aircraft turbulence of moderate or greater can significantly impact flight safety and is of the most concern to aviators. Moderate turbulence is classified as "unsecured objects are dislodged; occupants feel definite strains against seat belts and shoulder straps" (Schwartz 1996). Severe turbulence is defined as "occupants thrown violently against seat belts; momentary loss of aircraft control; unsecured objects are tossed about (Schwartz 1996). Table 1 relates turbulence intensities to a numerical value, which is reported in communication circuits and recorded in databases (Schwartz 1996). Using a turbulence diagnostic without first calibrating the threshold values for a given model and grid setup could lead to misleading forecasts of turbulence. The goal of this analysis is to

determine, for the chosen period of time, the most appropriate Ellrod TI2 index threshold with which to forecast moderate or greater turbulence forecasts. This was done through the use of a few basic accuracy and utility measurements. The index value chosen, after reviewing the outcome of the results from this research, is not meant to be a permanent value for use outside of this thesis. The value was chosen only for use in demonstrating the utility of the probabilistic forecast. The reader will recall from the Background chapter that TI index value is model-dependent and maybe seasonally dependent as well.

Value	**Intensity**
0	None
1	Light
2	Light-Moderate
3	Moderate
4	Moderate-Severe
5	Severe
6	Severe-Extreme
7	Extreme
9	Missing

Table 1. Numerical Value Assigned to PIREPS (after Table 1, Schwartz 1996)

2. Ellrod and Knapp's Approach

Ellrod and Knapp (1992) chose to verify TI1 and TI2 by examining four issues: event statistics, a Canadian validation study, threat scores, and frequency distributions. For event statistics, Ellrod and Knapp (1992) verified their two indices, which were used with different models (TI1-NCEP models and TI2-AFWA model) separately. They verified TI1 subjectively by comparing forecast CAT "events" with PIREPS during a six hour window (three hours on either side of the analysis time). They defined an "event" as "an area of index values greater than the threshold for the particular numerical model being evaluated." They required at least two reports of moderate or greater intensity turbulence inside the threshold index contour. If no reports were found within the threshold contour, the area was not verified. Ellrod and Knapp (1992) used reports from 20, 000 to 35,000 feet to verify forecasts created by the NCEP models of a 300 to 400 mb layer. They verified TI2 High Resolution Analysis System (HIRAS) output with manually derived Northern Hemispheric turbulence analyses produced by forecasters.

The areas produced by both the TI2 HIRAS and human subjective turbulence forecasts were compared. Areas with at least one-third overlap of the TI2 and forecaster derived areas were considered “hits.”

Ellrod and Knapp (1992) also examined a Canadian validation study, which evaluated the effectiveness of the Ellrod Index versus an index called the Empirical Index. They noted that the Ellrod Index had an overall success rate similar to the Empirical Index, but had a slightly better false-alarm rate (FAR as defined by Ellrod and Knapp 1992).

Since the probability of detection (POD) and FAR used by Ellrod and Knapp (1992) only provide some information about forecast system reliability, they examined the index using a measurement called the threat score. The sporadic nature of turbulence and the under-sampled nature of turbulence observations mean that the threat score is a better measurement to use than some other measurements.

Ellrod and Knapp (1992) also collected frequency distribution statistics for events by CAT intensity versus index values over grid points from the AVN model of the time. Using those statistics they were able to choose the best index value for detecting moderate or greater turbulence. Each model setup used in the Ellrod and Knapp (1992) study required different TI index threshold values for turbulence intensity. For example, the AFWA model required a threshold of TI 8 for MDT turbulence, 4 for NCEP’s NGM, and 2 for NCEP’s AVN. The different threshold levels for each model demonstrate the Ellrod Index’s variability between models.

3. Approach

To choose an appropriate threshold value for generating a probabilistic forecast, an objective process was implemented by which deterministic-based analyses (00 hour forecasts) of TI2 (Equation 6 - the second Ellrod turbulence index, which includes CVG) were created from 13 days in September 2005 (specifically 14-24 September and 29-30 September). These dates were chosen based on data availability. Thresholds of TI2 >1, 1.5, 2, 3, 4, 4.5, 5, 5.5, 6, 7, 8, and 9 were analyzed. The analyses were compared to PIREPS inside a six hour window (three hours on either side of the analysis time). For

this thesis, effort was made to take a similar but not identical approach to calibration as Ellrod and Knapp (1992).

Departing from Ellrod and Knapp's verification methods, this present study objectively determined verification statistics for the TI2 by using the following methods. The PIREP database was subdivided into two subsets. Those PIREPS with numerical intensity values of three or greater (see Table 1) were considered observed YES for moderate or greater turbulence and PIREPS with values less than three were considered observed NO for moderate or greater turbulence. Only PIREPS from 30, 000 to 39,000 feet were used to compare with turbulence forecasts of a layer from 30,000 to 39,000 feet. Turbulence forecasts (00 hour forecast) were created for each grid point over the entire globe. PIREPS were adjusted to the nearest grid point by rounding the latitude and longitude of the PIREP location to the nearest degree. Only grid points where PIREPS were reported were considered in the verification statistics. Grid points with forecasts and no PIREPS were not verified. Figure 6 is an example of how the PIREPS would look if overlaid onto the analysis of Ellrod TI>5. The circles represent null reports or reports of turbulence 2 or less from Table 1.

Ideally, PIREPS associated with thunderstorms would have been discarded, since the Ellrod turbulence diagnostic theoretically only identifies CAT and not convective related turbulence. This was not done in the determination of an index threshold, therefore thunderstorm contamination may have produced error in the verification results. An additional source of error will come as a result of automating the calculations for verification statistics. One single null report needed to be introduced into the PIREPS dataset on September 17, 2005 for computer programming reasons. This additional null report will impact the observed NO column (b or d from Table 2, depending on whether or not the event was forecast) of the contingency tables and any verification statistics using the particular observed NO value, but will likely only impact the values slightly since there are so many null reports.

Control MDT-SVR Turbulence (30 to 39 kft) TI Threshold >5 Date: 050914 Runtime: 00Z Forecast: f00hr

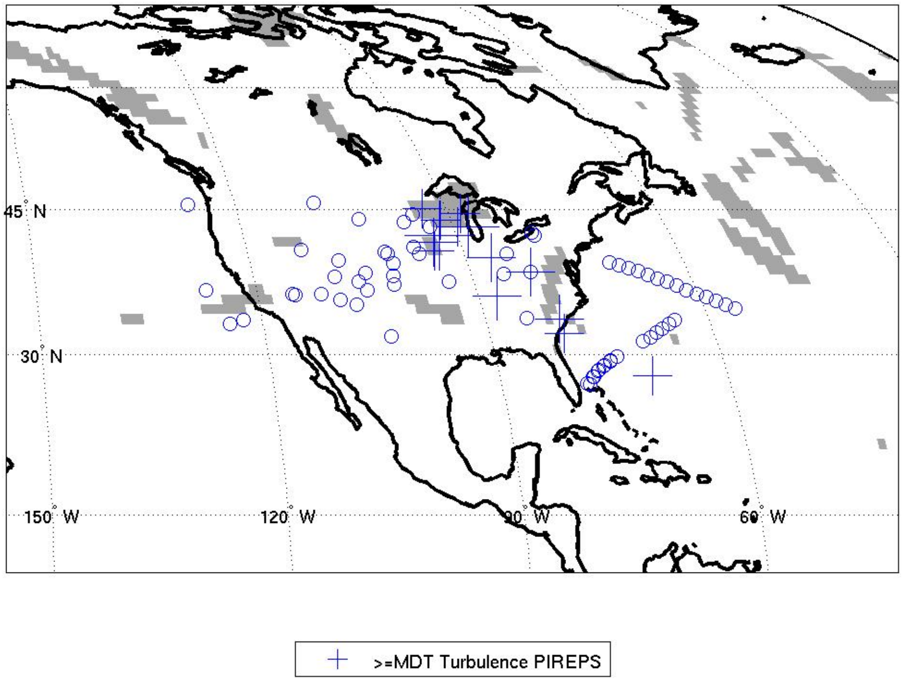


Figure 6. Example Plot of PIREPS overlaid with Ellrod TI2>5 shaded in gray (Circles are null reports).

4. Statistical Measures Used

A contingency table of absolute frequencies using (Wilks 2006) was setup for the results generated analyzing this issue (see Table 2). It is important to note that many different authors define statistics derived from contingency tables differently. For example, Ellrod and Knapp (1992) defined probability of detection (POD) and false-alarm ratio (FAR) differently than Wilks (2006). Ellrod and Knapp (1992) based their definitions from Weiss (1977). Using the Table 2, their equations would be defined as:

$$POD = \frac{a+d}{a+d+c} \tag{8}$$

$$FAR = \frac{b}{a+d+b}. \tag{9}$$

The previous definitions are different from the POD and FAR defined by Wilks (2006) and from the FAR defined by Zhu et al. (2002). Using the contingency table (Table 2), Wilks (2006) defined POD and FAR (from equations 7.7 and 7.8) as:

$$POD = \frac{a}{a+c} \tag{10}$$

$$FAR = \frac{b}{a+b} \quad . \tag{11}$$

Zhu et al. (2002) defines FAR as:

$$FAR = \frac{b}{b+d} \quad . \tag{12}$$

As one can see, definitions of accuracy measurements differ among authors, therefore the equations in Table 3 will serve as definitions for accuracy and utility measurements in this thesis.

	Observed		
Forecast		**Yes**	**No**
	Yes	a	b
	No	c	d

Table 2. Sample Contingency Table (after Fig 7.1, Wilks 2006).

The most basic accuracy measurement is the hit rate (HR) defined in Table 3. This measurement simply describes the proportion of correct forecasts when considering n forecasting occasions (Wilks 2006). The best possible hit rate is one and the worst is zero. The HR (eq. 17) will not be very helpful for this data analysis since there are a large number correct NO forecasts (reports of turbulence area a rare event). Another accuracy measure is the POD (eq. 14), which is “the likelihood that the event would be

forecast, given that it occurred” (Wilks 2006). A perfect forecast yields a one and a poor forecast yields a zero. If one is not concerned with the FAR (eq. 15), then the POD is a sufficient accuracy measurement. Most forecast users are interested in both the POD and FAR. A POD too low and a FAR too high are generally not acceptable. The FAR is simply the proportion of false alarms to total forecasts. To balance the POD and FAR other accuracy measurements and skill scores should be considered when analyzing a forecast system’s performance. The TS (eq. 16) is one such accuracy measurement. The TS is a good measurement for when there are a large number of correct NO forecasts, since it does not account for them. The TS is the proportion of correct YES forecasts to the “total number of occasions on which the event was forecast and/or observed” (Wilks 2006). Bias indicates whether the forecast system overforecasts (B>1) or underforecasts (B<1) a phenomena and is based on the total YES forecasts versus the total YES observations.

Another good measurement for skill is the Heidke Skill Score (HSS). The HSS uses the HR as the basic accuracy measure for the forecast system, but then compares the forecast system accuracy with the accuracy that could be achieved by random forecasts. Heidke Skill Scores of zero indicate the forecast system is equivalent to random forecasts, perfect forecasts receive a one, and a score less than zero implies the forecast system is worse than random forecasts (Wilks 2006).

The ZFAR and ZHR defined in Table 3 will be used to create a Relative Operating Characteristics (ROC) diagram plot. These plots help indicate whether or not a forecast system has the ability to distinguish between events and non-events (Zhu et al. 2002). The ROC area (ROCA) is used as a summary measure defined as the area between the point representing the system (ZFAR, ZHR), (0,0), and (1,1). Zhu et al. (2002) went further to describe the ROC area-based skill score (ROCS) as (from eq. 8, Zhu et al.):

$$ROCS = 2(ROCA - 0.5)\,. \tag{13}$$

The ROCS indicates the overall utility of the forecast system (Zhu et al. 2002). Larger values indicate more utility.

Equation Name and Source	Equation
Probability of Detection (Wilks 2006)	$POD = \frac{a}{a+c}$ (14)
False-alarm Ratio (Wilks 2006)	$FAR = \frac{b}{a+b}$ (15)
Threat Score (TS) (Wilks 2006)	$TS = \frac{a}{a+b+c}$ (16)
Hit Rate (HR) (adapted from Wilks 2006)	$HR = \frac{a+d}{n}$, where $n = a+b+c+d$ (17)
Bias (Wilks 2006)	$B = \frac{a+b}{a+c}$ (18)
Heidke Skill Score (HSS) (Wilks 2006)	$HSS = \frac{2(ad-bc)}{(a+c)(c+d)+(a+b)(b+d)}$ (19)
Zhu Hit Rate (ZHR) (Zhu et al. 2002, also known as POD by Wilks 2006)	$ZHR = \frac{a}{a+c}$ (20)
Zhu False-alarm Ratio (ZFAR) (Zhu et al. 2002)	$ZFAR = \frac{b}{b+d}$ (21)

Table 3. Table of Equations – Accuracy and Utility Measurements for Forecast Verification.

5. Results for Turbulence Index Threshold Calibration

Analysis of the performance of the Ellrod index (TI2) at several thresholds was conducted for a short period of time (14-24 and 29-30 September 2005). Overall, the

results suggest that a threshold near approximately TI2 > 5 indicates the most accuracy and utility for the ETFS setup (refer to Methodology chapter for ETFS setup details).

a. Accuracy Measurements

The POD and FAR (Table 3) indicate a decreasing POD and FAR with increasing TI2 threshold values (Figure 7). As the TI2 threshold increases, the geographic area of TI2 decreases. Since POD does not account for observed NO situations, the likelihood of detecting observed events decreases as the area decreases. The FAR decreases as a result of the decreasing TI2 area (i.e., decreasing sensitivity). These results do not provide a clear indication of which TI2 threshold is most reliable.

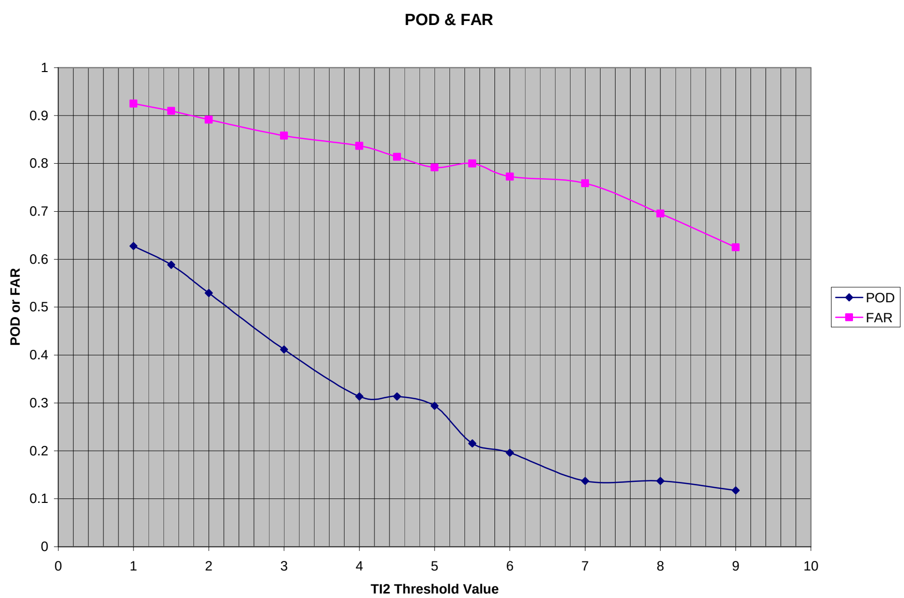


Figure 7. POD and FAR vs. TI2 Threshold Value.

The HR (Table 3) increases with TI2 threshold. The increase occurs as a result of increased number of observed NO cases, which increases the number of correct forecasts (Figure 8). That is, as the TI threshold coverage area decreases with increasing threshold, the number of observed NO/forecast NO cases increases. Because of the large number of null reports, the hit rate does not assess reliability well.

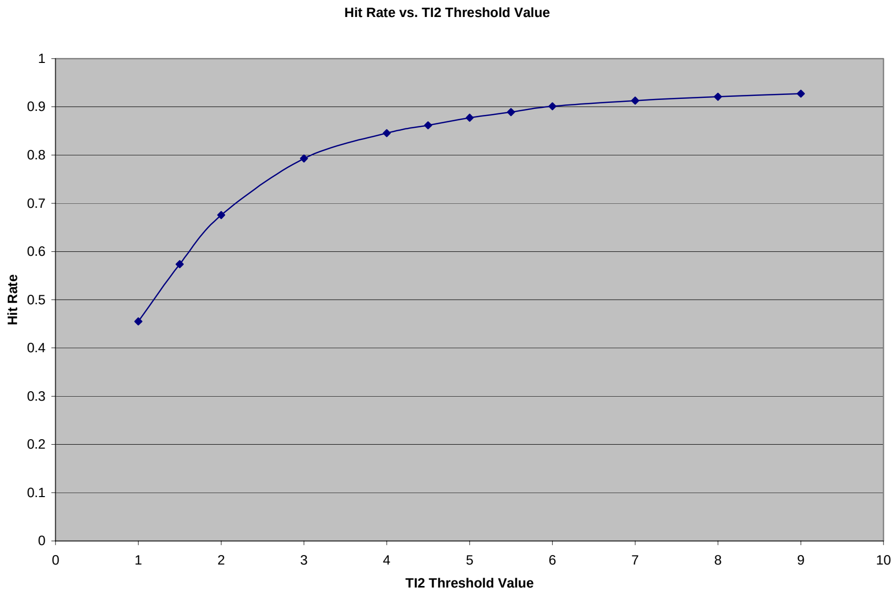


Figure 8. Hit Rate vs. TI2 Threshold Value

b. Utility Measurements

The TS, HSS, and Bias (Table 3) distinguished a best performing turbulence index threshold compared to the accuracy measurements in the preceding sub-section. A TI2 threshold of five has the highest TS and HSS (Figure 9). Therefore, an area inside the TI2 > 5 contour yields the highest TS and HSS. The TS is like a hit rate, when correct NO forecasts are removed from consideration (Wilks 2006). The positive HSS indicates that the forecast system is most skillful when using a threshold value of >5. Figure 10 is a plot of TS and Bias versus TI2 Threshold Value. The most unbiased TI2 threshold value is 5.5, with a Bias value of 1.078 (Refer to Appendix B). The Bias for TI2 threshold value 5 is 1.411. Both threshold values indicate a tendency to overforecast the event.

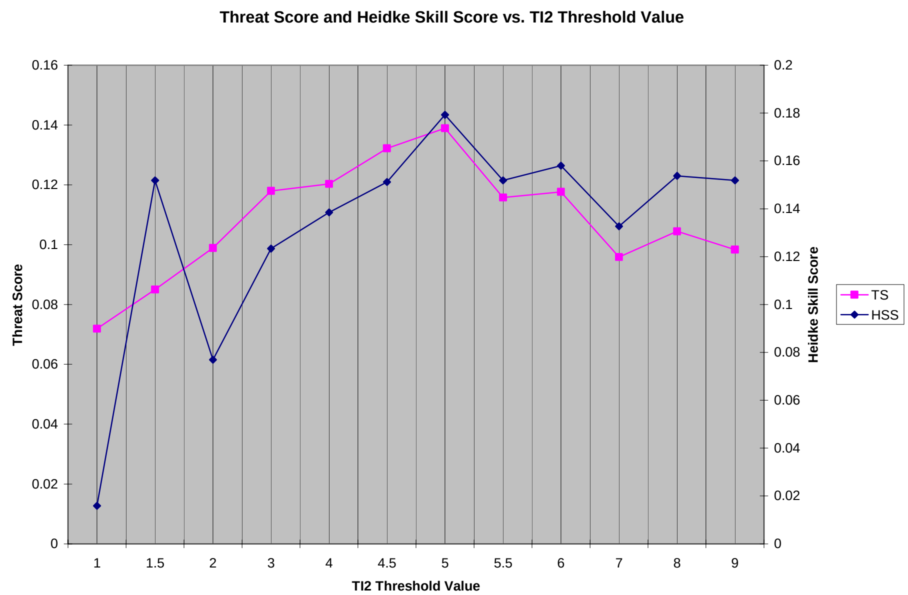


Figure 9. Threat Score and Heidke Skill Score vs. TI2 Threshold Value

Figure 10. Threat Score and Bias vs. TI2 Threshold Value

The largest ROCA occurs with TI2 thresholds three and five (Figure 11), therefore the forecast system set to TI2 thresholds of three and five exhibit the best ability to distinguish between conditions under which a certain event does or does not occur.

Values closest to the upper-left corner indicate more utility. Figure 12 is a plot of ROCA and ROCS. Higher values indicate more utility than those thresholds with lower values. It can be seen that there are two maxima near a TI2 threshold value 3 and 4.5 to 5.

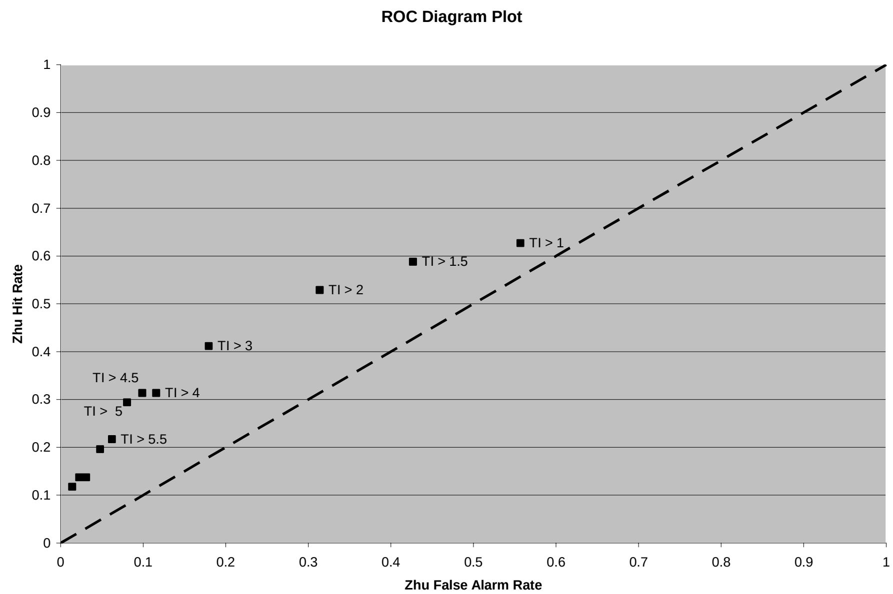


Figure 11. ROC Diagram Plot for Threshold Determination

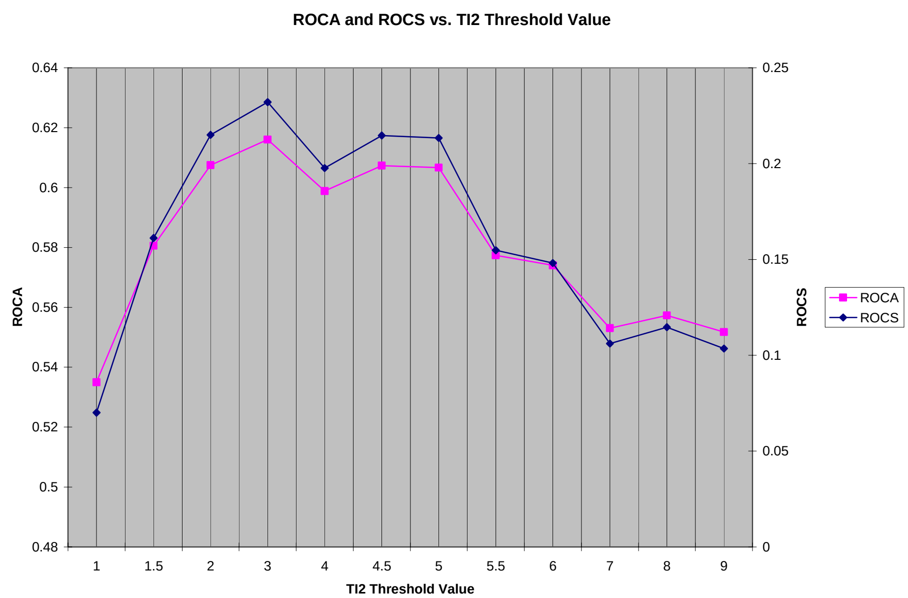


Figure 12. ROCA and ROCS versus TI2 Threshold Value Plot

c. Summary

As a result of the above analysis, a threshold of TI2 > 5 was chosen as the best TI2 threshold to use for the ETFS. First, TI2 > 5 had the highest TS and HSS score out of all of the thresholds. That means that TI2 > 5 forecasts the highest proportion of correct forecasts after correct NO forecasts are removed. In addition, it means that the threshold TI2 > 5 is the most skillful when compared to random forecasts as the reference forecast. While the Bias at TI2 > 5 is slightly above one (preferably Bias=1), it is relatively close compared to most other thresholds. Finally, Figure 12 illustrates that TI2 > 5 is a second maxima for being able to best distinguish between events and non-events.

D. GENERATING FORECAST PROBABILITY

Once the various ensemble members are generated from an EPS, there are several methods for generating forecast probability (FP) from those ensemble members. All examples in this sub-section will be explained with the understanding that a hypothetical EPS produces 10 ensemble members that provide turbulence measures (40 ensemble members were used in the actual experiment, however, 10 are used in this subsection for illustrative purposes).

Perhaps the most basic method for generating FP is what Eckel (1998, 2003) called the democratic voting method. With this method, each ensemble member gets an equal vote. For example, if a single threshold were used, such as the turbulence diagnostic TI > 4, the number of members that exceed this threshold would be divided by the total number of ensemble members to yield a FP. Unfortunately, the democratic voting method does not properly account for FP. The democratic voting method does not account for a small amount of probability in partial bins. This weakness eliminates important forecast probability detail. Figure 13a illustrates the democratic voting method. Using the assumptions in the example, the democratic voting method generates a FP = 6/10 = 0.6.

If the uniform ranks method is performed on the same set of ensemble members, the FP = 0.6061. The uniform ranks method assumes that there is a uniform probability distribution of the ensemble members, with each member being equally likely. With the uniform ranks method, an additional fraction of a rank probability bin must be taken into

account (see Figure 13b). This is done by linearly interpolating the distance between the threshold and the ensemble values on either side. The probability of that additional bin fraction is given by Eckel (1998; 2003):

$$P(T < V < x_i) = \left(\frac{x_{i+1} - T}{x_{i+1} - x_i}\right) RP_{i+1} \,. \qquad (26)$$

Here, T is the threshold value, V is the verification value, x_i is the value of the ensemble member with rank i, x_{i+1} is the value of the ensemble member of i+1, and RP_{i+1} is the amount of the probability of the verification rank i+1 (that is $RP_{i+1}=1/(n+1)$). For the uniform ranks method the following equation is used for FP:

$$FP = P(T < V < x_i) + (\# of members > T) / (n+1)) \,, \qquad (27)$$

where n is the number of ensemble members. Eckel (1998, 2003) adapted the uniform ranks method from Hamill and Colucci (1997).

Eckel (2003) noted that the democratic voting method pushes FP to extreme values, such that high FP is overestimated and low FP is underestimated. Further, he demonstrated that low sampling exaggerates the problem. Weighted ranks (or calibrated) is a better method for generating FP than the uniform ranks method (Eckel 1998; 2003). In this research, calibration is not explicitly defined, so only the uniform ranks method has been employed.

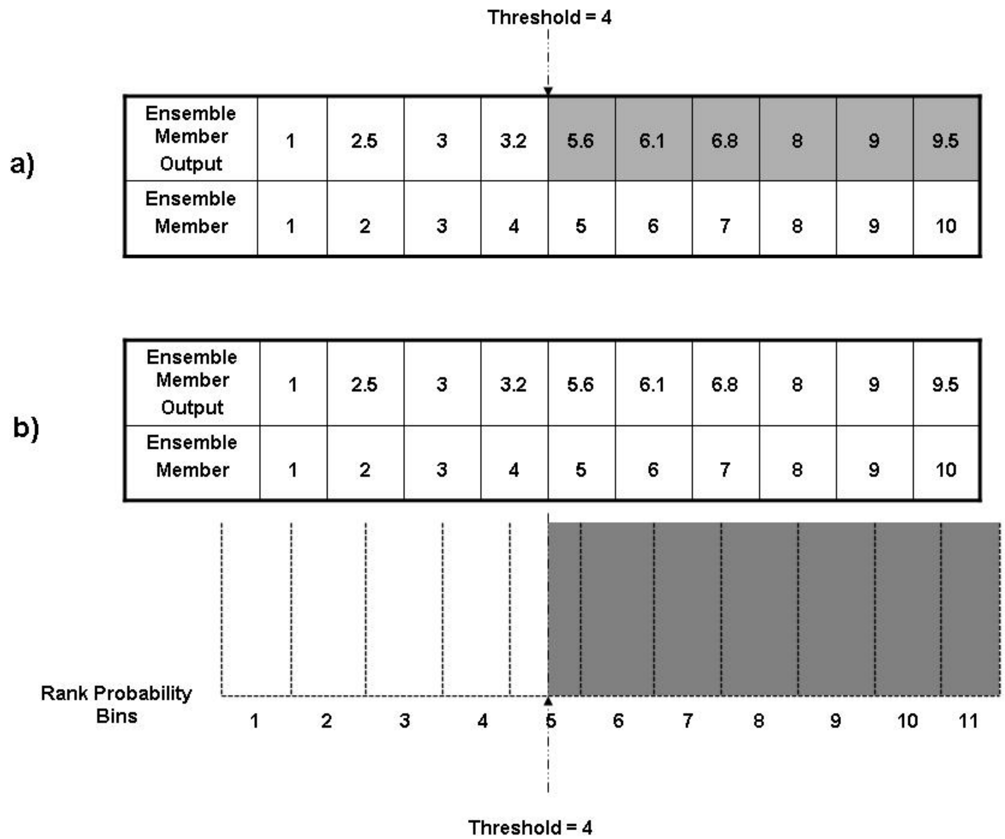


Figure 13. Calculating Forecast Probability: a) Democratic Voting Method and b) Uniform Ranks Method (after Figure 38 & 39, Eckel 2003)

Another advantage of the uniform ranks method over the democratic voting method occurs with extreme forecast probability (if all members are above or below the threshold). For example, if all members exceed the threshold, say 10 out of 10 ensemble members have a TI2>4, then the democratic voting method would yield a forecast probability of 100% (see Figure 15). Alternatively, if none of the members exceed the given threshold, the democratic voting method would yield a FP of 0%. Eckel (1998, 2003) explains that the Gumbel distribution in (Wilks 2006) can be used to characterize extreme-value data. In our case, the Gumbel CDF is used to calculate low probability situations (see Figure 14). The Gumbel distribution is best used for right tail situations, since the distribution is skewed to the right. The Gumbel CDF (Wilks 2006) is:

$$F(x) = \exp\left\{-\exp\left\{-\frac{(x-\xi)}{\beta}\right\}\right\} \tag{28}$$

The estimation equations for the Gumbel distribution are (Wilks 2006):

$$\hat{\beta} = \frac{s\sqrt{6}}{\pi} \tag{29}$$

$$\hat{\xi} = \bar{x} - \gamma\hat{\beta} \tag{30}$$

$\gamma = 0.57721$. (Euler's Constant)

To find the FP of a low probability situation, the following equation, adapted from Eckel (1998; 2003), is used:

$$P(T < V) = \left(\frac{1 - F(T)}{1 - F(x_{10})}\right) RP_{11}. \tag{31}$$

Note that in Figure 14 there is no ensemble value to the right of the last bin to use to calculate the probability. The Gumbel distribution is assumed to find a theoretical value on the right. Equation 31 is very similar to equation 26. The estimation equations must be evaluated using the sample data available. For the example in Figure 14, using the method of moments $\beta = 0.681511$ and $\zeta = 2.686625$, where $\bar{x} = 3.08$ and $s = 0.874071$. The Gumbel CDF probability for the threshold is $F(T)$ and $F(x_{10})$ is Gumbel CDF probability for the last ensemble member value. Finally, $P(T<V) = 0.079387$. Assuming a probability distribution function for extreme value produces more realistic forecast probability, as in this case where FP = 7.9% as opposed to 0% using democratic voting method.

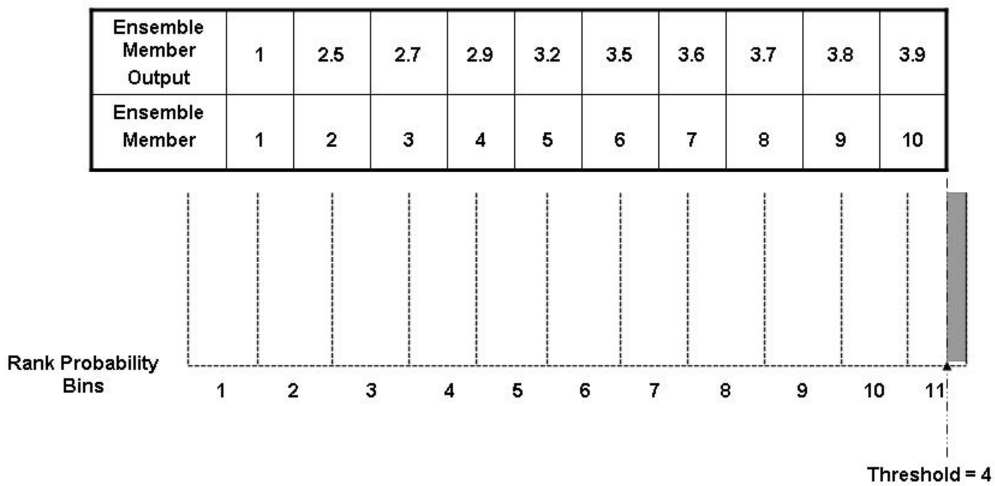


Figure 14. Calculating Forecast Probability with Low Probability Situations – Using the Gumbel CDF (after Figure 38 & 39, Eckel 2003)

Extreme high forecast probability situations (all ensemble members higher than threshold) must also be addressed by assuming a theoretical distribution (see Figure 15). A reverse Gumbel would work in our case, since the Ellrod index does produce some negative values. However, negative values tend to produce little to no turbulence (Ellrod and Knapp 1992). Therefore, the following distribution (from Eckel 2003) was chosen as for the extreme high probability situations:

$$P(V > T) = \left(1 - \left(\frac{T}{x_1}\right)^3\right)\frac{1}{n+1} \ . \qquad (32)$$

In Figure 15, the democratic voting method would generate a forecast probability of 100%. If one were to assume the PDF given by equation 32, the forecast probability

would be 0.953455 or 95.3%. An additional 10/11 must be added to the value generated by equation 32 for final probability in a high probability situation.

Ensemble Member Output	5	5.6	8	9	9.5	9.8	12	13	13.5	13.6
Ensemble Member	1	2	3	4	5	6	7	8	9	10

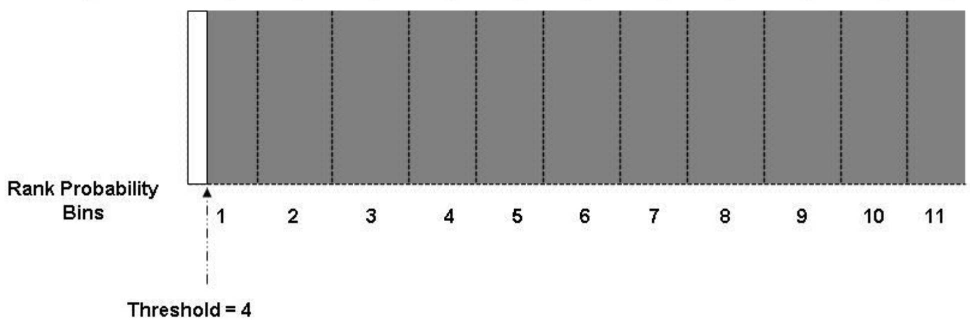


Figure 15. Calculating Forecast Probability with High Probability Situations - Using Equation 14 (after Figure 38 & 39, Eckel 2003).

E. GENERATING 40 ENSEMBLE MEMBERS USING LAGGED-AVERAGE FORECASTING

The NCEP GFS ensemble model generates 10 ensemble members at 00Z, 06Z, 12Z and 18Z and one member control forecast. The ten members are generated by defining five sets of positive and negative perturbations of the model control. To increase the number of ensemble members available to the ETFS, lagged-average forecasting was employed to generate a total of 40 ensemble members. The control run was not included as one of the 40 members of the ETFS. The ten positive and negative perturbation members were included from 00Z, 06Z, 12Z, and 18Z to create forecasts ranging from 00 hr to 72 hour forecasts. Lagged-average forecasting has been demonstrated as a viable and possibly better method than Monte Carlo methods for generating additional ensemble members (Kalnay 2003). Lagged-average forecasting involves using forecasts from previous model runs to increase the number of ensemble

members in an EPS. The older forecasts (forecasts from previous model runs) should be weighted according to their expected error. The statistics required to estimate weights according to the "age" of the ensemble are difficult to obtain (Kalnay 2003). A strong advantage of lagged-average forecasting over other methods, is that the forecasts are already available for use. No additional computer time is needed for ensemble members introduced through this method. Unfortunately, lagged-average forecasting without appropriate weighting factors may lead the old forecasts to negatively taint the ensemble average. Weighting of forecasts by "age" was not done for this research, due to the difficulty of obtaining appropriate weights for older forecasts. Each member will be weighted equally. Refer to the tables in Appendix B for details on how the lagged-average forecasting was actually setup for this research.

F. FINAL PRODUCTS

The final products of the ETFS are global (1.0 gridded) forecasts of the probability of occurrence of moderate or greater aircraft-scale turbulence produced in five different layers, based on U.S. standard atmosphere heights and pressures. Table 4 defines the levels generated by the ETFS. Figure 5 is an example of a Layer-1 24-h forecast. Forecasts for 06, 12, 18, 24, 30, 36, 48, 54, and 72 hours were made for each layer. Recall, Figure 5 is an example of a probabilistic turbulence forecast product from the ETFS. Figure 5 represents the forecast probability of moderate to severe turbulence for the layer 30,000 ft to 39,000 feet. Warm colors represent a high probability of moderate to severe turbulence for the given layer represented in the figure. Cool colors represent lower probabilities of moderate to severe turbulence. These prototype forecast products are now utilized to demonstrate the value of probabilistic forecasts over traditional deterministic forecasts (Chapter V) and to illustrate the integration of probabilistic forecast information with Air Force decision-making (Chapter VI).

Layer	Pressure Height (mb)	Geopotential Height (kft)
1	200 to 300	30 to 39
2	250 to 350	26.5 to 34
3	300 to 400	23.5 to 30
4	350 to 450	21 to 26.5
5	400 to 500	18.5 to 23.5

Table 4. Table of Forecast Probability Vertical Layers Generated by the ETFS

THIS PAGE INTENTIONALLY LEFT BLANK

V. COMPARISON OF PROBABILISTIC VERSUS DETERMINISTIC AIRCRAFT-SCALE TURBULENCE FORECAST SYSTEMS

A. OVERVIEW

This section will demonstrate the economic advantage of using probabilistic forecasts versus deterministic forecasts of aircraft-scale turbulence. The analysis was conducted by performing cost-loss analysis similar to that which was conducted by Zhu et al. (2002) and also similar to that discussed by Wilks (2006). As discussed in the Chapter II, Zhu et al. (2002) clearly demonstrated the economic advantage of using forecast probability information versus deterministic forecast information; however, instead of using 500 hPa as the meteorological variable, the Ellrod Turbulence Index TI2 was chosen for thesis context.

Zhu et al. (2002) notes that forecast users "either do, or do not take action" regarding their response to a weather forecast. Either way, the forecast user's action or inaction leads to an expense related to protection, loss or no expense at all. Table 5 is a simple contingency table that relates the expense to hit, misses, false-alarms, and correct rejection. The table here uses similar variables as Zhu et al. (2002).

		Observed	
Forecast		**Yes**	**No**
	Yes	Hit (h) Mitigated loss $(C + L_u)$	False Alarm (f) Cost (*C*)
	No	Miss (m) Loss $(L = L_p + L_u)$ L_u ignored	Correct Rejection (c) No Cost (*N*)

Table 5. Simple Contingency Table Relating Expenses of Action or Inaction. L_u is an unprotectable loss and L_p is a protectable loss (after Table 1, Zhu et al. 2002)

B. COST/LOSS ANALYSIS

The cost-loss analysis methods used in this thesis are similar to those methods used in (Zhu et al. 2002; Murphy et al. 1985; and Wilks 2006). Table 5 is a simple table that accounts for costs and losses accumulated due to forecast user's action or inaction. The cost, C, refers to the cost incurred for protection, L_u refers to an unprotectable loss, L_p refers to a loss that can be protected against, and N refers to no cost. When there is a hit, the user's protection prevents a loss (L) from occurring, but incurs a cost of protecting (C) and an unprotectable cost (L_u). Unprotectable loss, L_u, will be ignored, in this thesis. For a false alarm, the user only incurs a cost of protection (C). For a miss, the user incurs the cost of a loss (L), which includes the protectable loss and unprotectable loss. And for a correct rejection, no cost is incurred (N). The assumption is that the cost of protection is less than a loss (i.e. $C<L$). A C/L user will react to forecast information when the probability forecast of an event occurring exceeds their particular C-L ratio. Meaningful C-L ratios are bounded by zero and one (i.e. $0<C/L<1$) (Wilks 2006). If the C-L ratio were not bounded by zero and one, the protective action offers no potential

advantages for values beyond the bounds. Further complexity can be added to the table to address more sophisticated requirements.

Ultimately, the goal of cost-loss analysis is to determine the economic value of one forecast system over another to see which forecast system provides the best utility for a user. Typically this is done by first calculating the expected expense of each forecast system, E_f, the expected expense using climate alone as a forecasting tool, E_c, and the expected expense of using a perfect forecast system, E_p. Expected expense is the "probability-weighted average costs and losses" (Wilks 2006). Once these values have been determined, the economic value of a forecast system can be calculated for each system. The equations for expected expense for a forecast system and a perfect forecast system can be found in Table 6. The expected expense for using climate alone is defined as (adapted from eq. 2, Zhu et al. 2002):

$$E_c = oL_u + Min\lfloor oL_p, C \rfloor \tag{33}$$

Unfortunately, a good turbulence climatology is a luxury not afforded for this thesis work. Therefore, where the expected expense of the climate, E_c, is typically used, another baseline will be used. The new baseline is based on never protecting. For example, an aviator would simply ignore forecasts of moderate or greater turbulence. Therefore, for each occurrence of a moderate or greater turbulence event (Observed YES, y), the aviator would incur a loss, L. Realistically, there is not a good climatology for aircraft-scale turbulence, so never protecting would be a viable option as a baseline expense to a user. The new baseline expected expense is defined as:

$$E_b = yL\,. \tag{34}$$

For a forecast system to be useful, its expected expense should be less than the baseline expected expense and therefore have more relative economic value.

Equation Name and Source	Equation
Expected Expense of Forecast System (adapted from eq. 1, Zhu et al. 2002)	$E_f = h(C) + fC + m(L_p)$ (35)
Expected Expense Using No-Protection Baseline (instead of Climate)	$E_b = yL$ (36)
Expected Expense of a Perfect Forecast System (adapted from eq. 3, Zhu et al. 2002)	$E_p = o(C)$ (37)
Economic Value (adapted from eq. 4, Zhu et al. 2002)	$V = \frac{E_b - E_f}{E_b - E_p}$ (38)

Table 6. Table of Equations for Cost/Loss Analysis

Relative economic value (equation 38) is a value that relates the expected expense for a baseline measurement, the forecast system and a perfect forecast system. A value of 1 is the maximum value, which represents a perfect forecast system. A value of zero indicates that the forecast system is no more valuable than using baseline alone (which would be never protecting). Finally, a negative value indicates that the forecast system actually costs more money than never protecting. Equation 38 is a modified version of the economic value equation used by Zhu et al. (2002). The results using the new equation appear slightly different than the results in Zhu et al. (2002).

C. EXPERIMENTAL SETUP

To only test the hypothesis that making decisions with probabilistic forecast information over traditional deterministic forecast information is better economically, several factors needed to be addressed and assumptions made. First, there was a need to eliminate various problems associated with turbulence verification. Since PIREPS alone

may not have provided enough valid observations to support the analysis, a single negative perturbation member's analysis (00hr forecasts) from a future model run is used as truth. For example, a 48-hour forecast generated on September 14, 2005 would match up with an analysis on September 16, 2005. By using the analysis as a truth, cost-loss analysis was conducted at every grid point. Using the analysis as truth probably artificially produces better or worse verification results than could be realistically obtained using existing turbulence observations systems. Therefore, any comparisons of verification results using analysis as truth versus using an observation system may not be a fair comparison. Secondly, the Ellrod Turbulence Index TI2 was used for turbulence calculations for the probabilistic forecast system, the deterministic forecast system, and the analysis (truth). For purposes of studying Objective 2 only, it is assumed that the Ellrod Turbulence Index forecasts turbulence accurately. The limitations of the Ellrod Turbulence Index are understood and were outlined in the background section, so they will not be addressed here. This experiment was setup only to test the effect of using probabilistic forecast information versus deterministic forecast data for aircraft-scale turbulence using cost-loss analysis and economic value measurements.

Truly, the Ellrod Index is nothing more than a diagnostic applied during pos-processing, so the results gained in this section should be consistent with the results generated reported in Zhu et al. (2002). For robustness, four case studies were created to examine the second objective. Table 7 relates important information about each case study. Due to data availability, only forecasts of 24 and 48 hours were analyzed for the September case studies. An additional 72 hour forecast was analyzed for the November case studies.

Only one turbulence layer was used for the calculations (Level 1 – 200 to 300 mb). Each grid point was treated as a separate forecast opportunity. Contingency table data were collected for each day during the case study time period. The expected expense was calculated for the ETFS (System A - probabilistic) and a single ensemble member (System B - deterministic) forecast system. The expected expense for a perfect forecast system and not protecting were also calculated for each day. The mean (average) expected expense for each system or baseline was calculated for all the days in the time period of the case study for C-L ratio's 0.05-0.95, every 5%. Finally, the

economic value of both System A and B were calculated from their respective mean expected expense for the given forecast period.

Case Study	Time Period	Geographic Coverage	Event Opportunities per forecast (# of grid points)
1	14-22 September 05	Global	65,160
2	14-22 September 05	30N to 55N	9,360
3	1-13 November 05	Global	65,160
4	1-13 November 05	30N to 55N	9,360

Table 7. Case Studies

D. RESULTS OF COST/LOSS ANALYSIS

1. Overview

To examine Objective 2, three types of plots were created for each case study; i) a plot of expected expense of the forecast systems (System A – probabilistic and System B - deterministic) vs. C-L ratio; ii) relative economic value of the forecast systems vs. C-L ratio; and iii) a ROC plot for each forecast system.

Hypothesis tests were conducted to determine if the difference between means of the two independent samples of expected expense values for each forecast system were statistically significant. The population standard deviations were unknown but treated equally. A two-sample student's t test (Anderson and Finn 1996) was used and defined as:

$$t_{stat} = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (39)$$

with f degrees of freedom defined as

$$f = n_1 + n_2 - 2. \qquad (40)$$

S_p is defined as the pooled standard deviation for both samples. A two-tailed hypothesis test (α=.05) was conducted with the following null and alternative hypotheses:

H_0: Mean E_f of System A (Probabilistic) = Mean E_f of System B (Deterministic)

and

H_1: Mean E_f of System A (Probabilistic) ≠ Mean E_f of System B (Deterministic),

respectively. Detailed hypothesis testing results can be found in Appendix C. In Appendix C, note that for some C-L ratios the difference in means are not statistically significant, meaning that we cannot reject the null hypothesis that for some C-L ratio there may not be any measurable value of one forecast system over another.

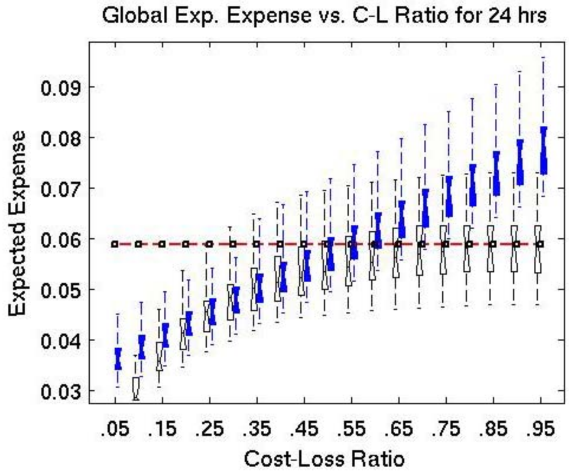
For each case, the construction of the economic value vs. C-L ratio figures within this thesis differ from that of Zhu et al. (2002) for extreme low probability events. The difference occurs as a result of the separate methods by which forecast probabilities were calculated. Recall for the ETFS a probability distribution based on Eckel (1998, 2003) is used to determine 'far right' values in extreme low probability forecasts in the uniform ranks method. This apparent difference is visible in the economic value figures at low C-L ratios. Zhu et al. (2002) notes that their low negative values for the ensemble at a very low C-L ratio are due to the small size of the ensemble. It should also be noted that the figures for each case study and those in Zhu et al. (2002) do not include C-L ratio end points 0.0 and 1.0. A C-L ratio of zero would imply that it costs nothing to protect. Therefore, a rational user would always protect and never incur a loss. On the other hand, a C-L ratio 1.0 or higher would imply that it costs as much or more to protect than the expense incurred due to a total loss. A rational forecast user would never protect if their C-L ratio were greater than or equal to 1.0. The previous statements about the extreme C-L ratios remain true for both deterministic and probabilistic forecast systems. Only values between, but not including 0.0 and 1.0 are meaningful in this C-L analysis.

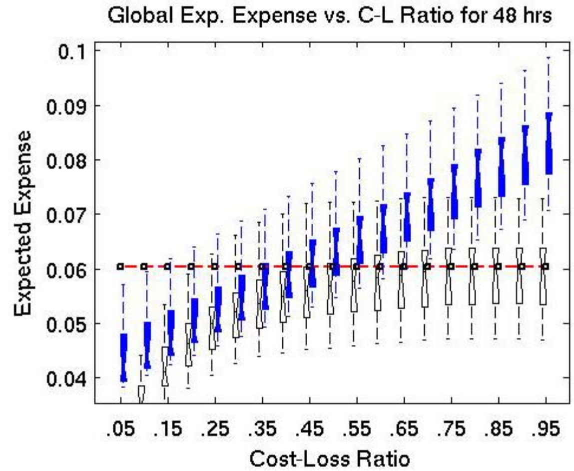
2. Case 1

At 24 hours the difference in means of expected expenses for System A (probabilistic) and System B (deterministic) are not statistically significant for C-L ratios of 0.15 to 0.6 (Appendix C). This is evident in Figure 16a as the expected expense values for System A and System B are nearly equal for those C-L ratios. Box and whisker plots are used to visualize the expected expense values for each forecast system. The rectangular portion of the box plot represents the middle 50% of the expected expense values for the particular time period. The whiskers are defined by 1.5 times the range of the middle 50% of the data. The line going through the box represents the median. Outliers are plotted as asterisks outside of whisker plots. At 48 hours the null hypothesis cannot be rejected at C-L ratios 0.15 to 0.5 (Appendix C). As anticipated, by 48 hours the expected expense for System A is less expensive relative to System B for more C-L ratios than it was for 24 hour forecasts. Furthermore, the expected expense values for System A differ markedly from System B for C-L ratios greater than 0.5 (Figure 16b).

For both 24-h (Figure 17a) and 48-h (Figure 17b) forecasts, there is little relative economic value of System A over system B for C-L ratios (less than 0.15). However, for many forecast users with high and low C-L ratios, System A has significantly more economic value than System B. At both forecast times, there is a marked difference in the economic values at C-L ratios higher than 0.6.

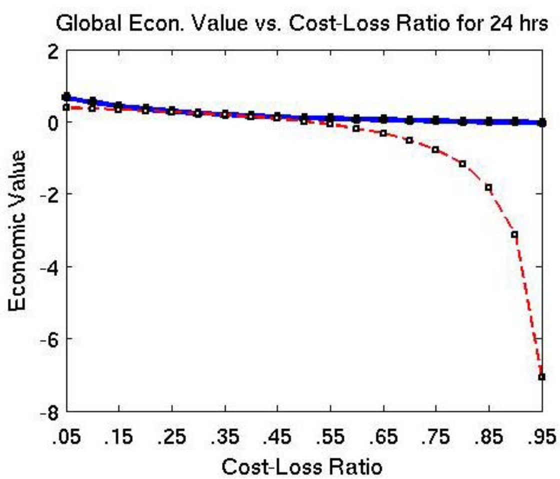
The ROC diagrams (Figure 18) demonstrate a clear advantage of probabilistic forecasts over deterministic forecasts. Because probabilistic forecast systems provide multiple decision levels, it is possible to plot a line that represents multiple comparisons of hit rate to false alarms (ZHR and ZFAR, as defined in Table 3). Deterministic forecasts, which only provide one decision-level are represented by the respective value for each forecast interval. The continuous nature of probabilistic forecasts increases the ROCA. Note that to calculate ROCA for a single deterministic forecast one would draw a triangle between vertices (0,0), the deterministic point, and (1,1). A triangle defined by those vertices would yield a smaller ROCA than the ROCA under the ROC curve for the probabilistic forecast system. It is important to note that ROC diagram plots indicate the potential skill of a forecast system that would only be achieved if the forecasts were correctly calibrated (Wilks 2006).

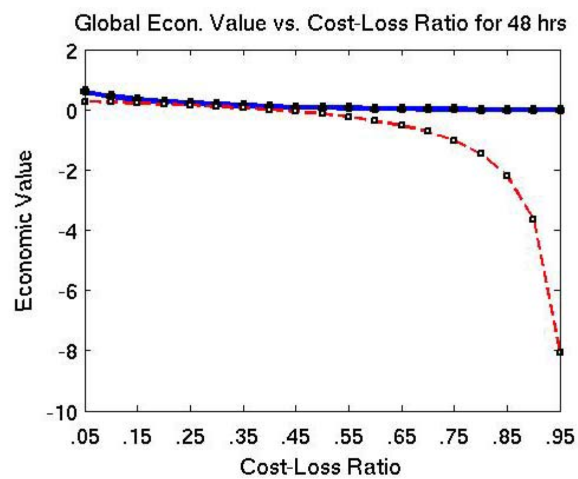




a) b)

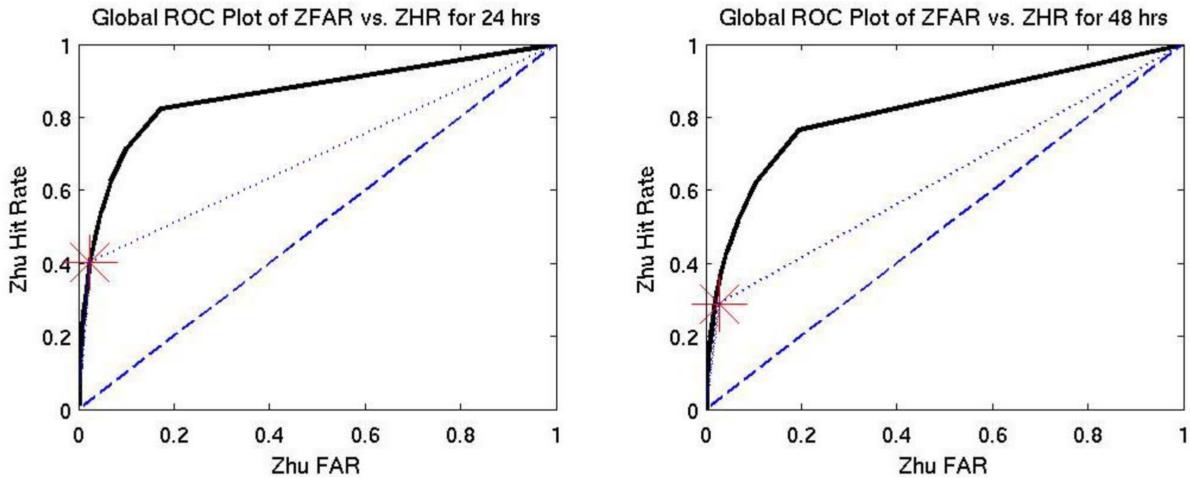
Figure 16. The expected expense value plots for Case 1 at (a) 24 hour and (b) 48 hours. The probabilistic model (System A) is defined by the unshaded box plots, the deterministic model (System B) is defined by the blue shaded box plots, and the median expected expense of not protecting is defined by the red dashed line.





a) b)

Figure 17. Economic value vs. C-L ratio plots for Case 1 of the probabilistic model (System A -blue solid line) and the deterministic model (System B – red dashed line) for (a) 24 hour and (b) 48 hour forecasts.



a) b)

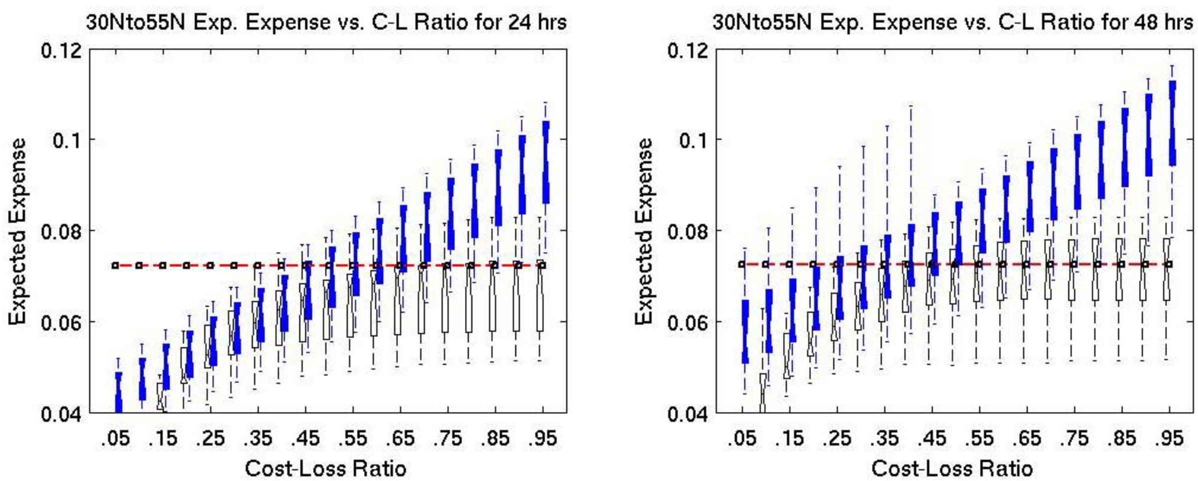
Figure 18. The ROC diagrams for Case 1, at (a) 24 hour and (b) 48 hour. The probabilistic system A is defined by the solid black line and the deterministic system is represented by the red star.

3. Case 2

Case Study 2 was limited to 30 N to 55N, to focus on the influences from the polar-front jet stream. This case study includes the same timeframe as Case 1. Hypothesis testing revealed that the expected expense from System A and System B may be the same for or C-L ratios 0.2 to 0.55 for 24 hour forecasts (Appendix C). For 48 hour forecasts, there is likely no difference in forecast system expected expenses for C-L ratios 0.15 to 0.65. Exactly why the system expected expenses are not clearly different at more C-L ratios is unknown. Figure 19 illustrates the expected expense for each forecast system at different C-L ratios for Case 2. As in Case 1, for C-L ratios where the null hypothesis cannot be rejected, the expected expense for System A and System B at 24 (Figure 19a) and 48 (Figure 19b) hours are very similar.

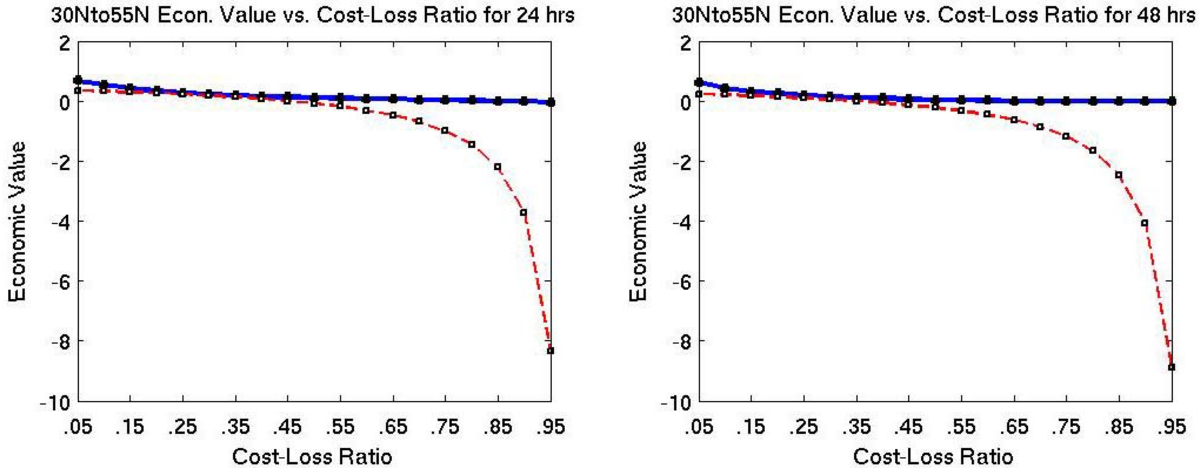
At 24 hours (Figure 20a) and 48 hours (Figure 20b), there is little relative economic value of System A over system B for C-L ratios 0.20-0.55. However, for many forecast users with high and low C-L ratios, System A has significantly more economic value than System B. Unlike Case 1 where increased forecast time showed an increase number of possible forecast users (with respect to C-L ratio), it cannot be said with certainty that more forecast users (in terms of C-L ratios) will benefit at 48 hours than 24

hours. Clearly, System A has more relative economic value than System B at C/L ratios higher than 0.65 and less than 0.15 at both 24 and 48 hours. Figure 21 demonstrates the inherent increase in utility that probabilistic forecasts have over deterministic forecasts. Probabilistic forecast systems have more utility because more forecast users can take advantage of the multiple decision levels.



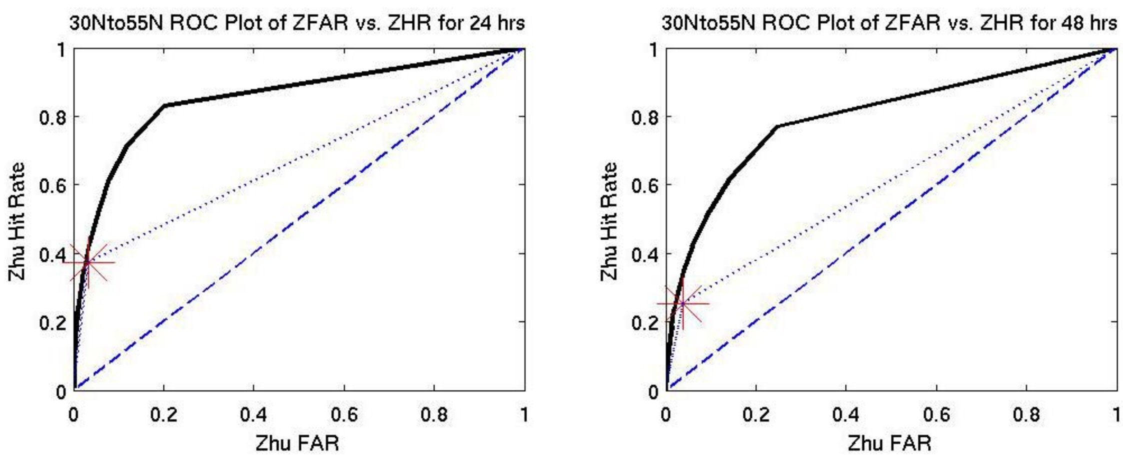
a) b)

Figure 19. The expected expense value plots for Case 2 at (a) 24 hour and (b) 48 hours. The probabilistic model (System A) is defined by the unshaded box plots, the deterministic model (System B) is defined by the blue shaded box plots, and the median expected expense of not protecting is defined by the red dashed line.



a) b)

Figure 20. Economic value vs. C-L ratio plots for Case 2 of the probabilistic model (System A -blue solid line) and the deterministic model (System B – red dashed line) for (a) 24 hour and (b) 48 hour forecasts.



a) b)

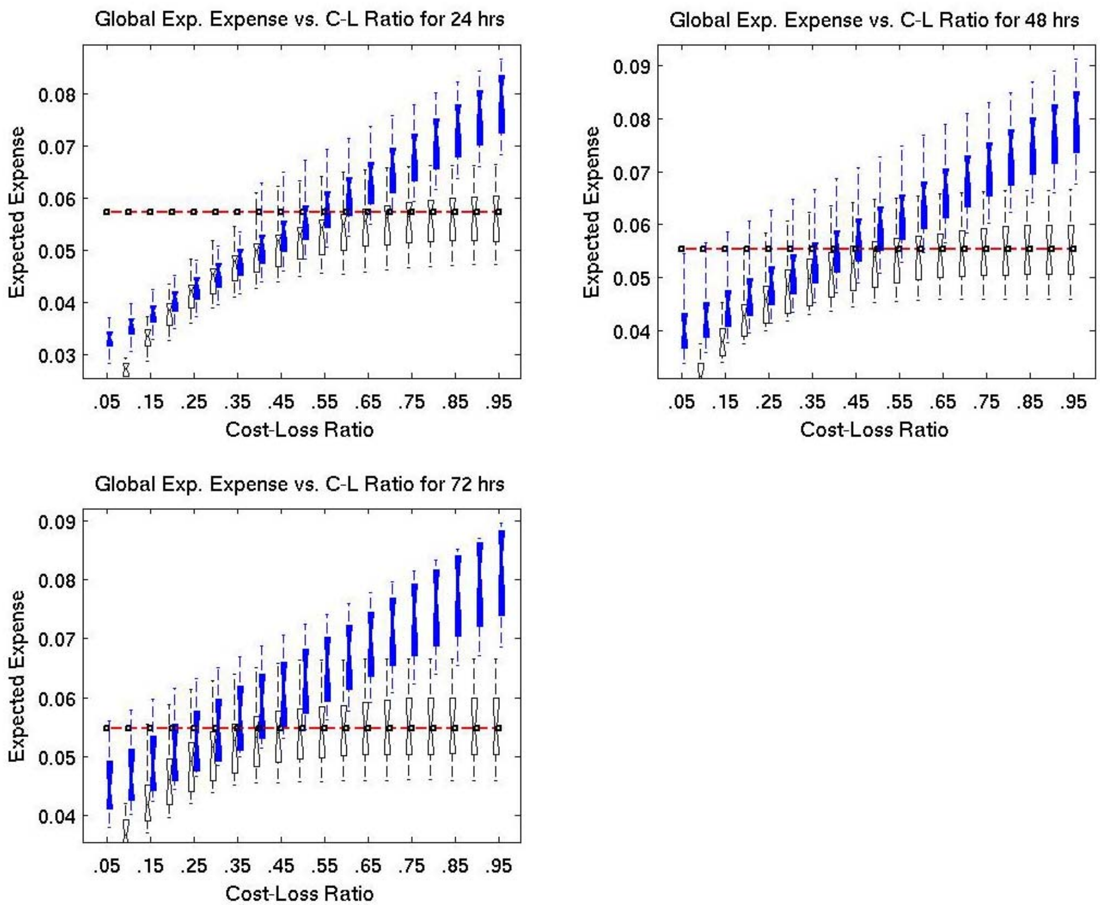
Figure 21. The ROC diagrams for Case 2, at (a) 24 hour and (b) 48 hour. The probabilistic system A is defined by the solid black line and the deterministic system is represented by the red star.

4. Case 3

Case 3 and Case 4 include an analysis of a 72-h forecast and consist of a November dataset (refer to Table 7). At 24 hours, the difference in the mean of expected expenses for System A (probabilistic) and System B (deterministic) are not statistically significant for C-L ratios of 0.2 to 0.5. At 48 hours, the null hypothesis cannot be rejected for C-L ratios 0.2 to 0.4. At 72 hours, the null hypothesis cannot be rejected for C-L ratios 0.25 to 0.3. As anticipated, by 72 hours the expected expense for System A is less expensive relative to System B for nearly all forecast users classified by C-L ratio. Figure 22 illustrates the expected expense for each forecast system at different C-L ratios for Case 3. Close examination reveals that the expected expense for System A and System B are very similar for C-L ratios where the null hypothesis cannot be rejected.

Figure 23 shows how that there is little relative economic value of System A over system B for some C-L ratios (i.e. 24 hour forecast C-L ratios of 0.2-0.5). However, for many forecast users with high and low C-L ratios, System A has significantly more economic value than System B. By 48 hours, System A begins to break away from System B. By 72 hours, System A has more relative economic value than System B for most forecast users based on C-L ratios. Figure 24 demonstrates the inherent increase in utility that probabilistic forecasts have over deterministic forecasts. Probabilistic forecast systems have more utility because more forecast users can take advantage of the multiple decision levels.

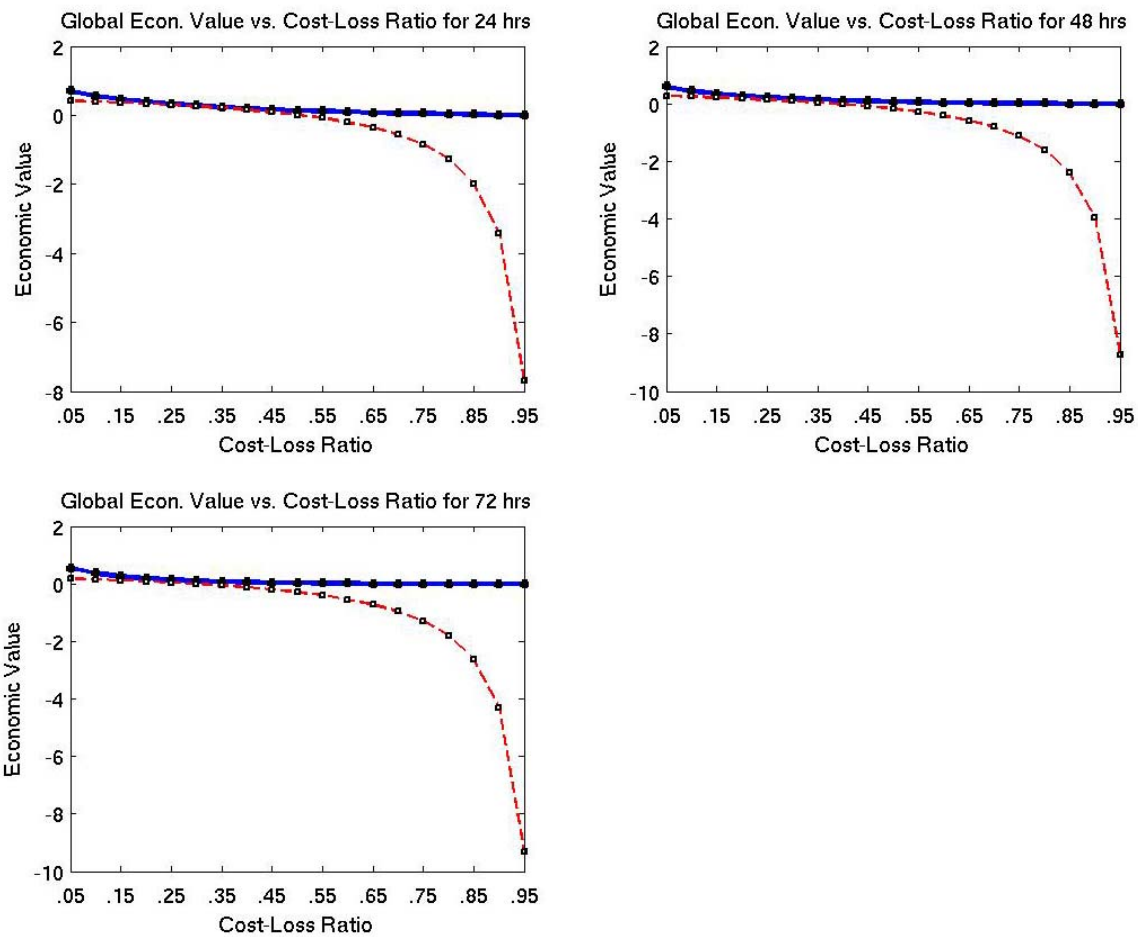
a) b)



c)

Figure 22. The expected expense value plots for Case 3 at (a) 24 hour, (b) 48 hours, and (c) 72 hours. The probabilistic model (System A) is defined by the unshaded box plots, the deterministic model (System B) is defined by the blue shaded box plots, and the median expected expense of not protecting is defined by the red dashed line.

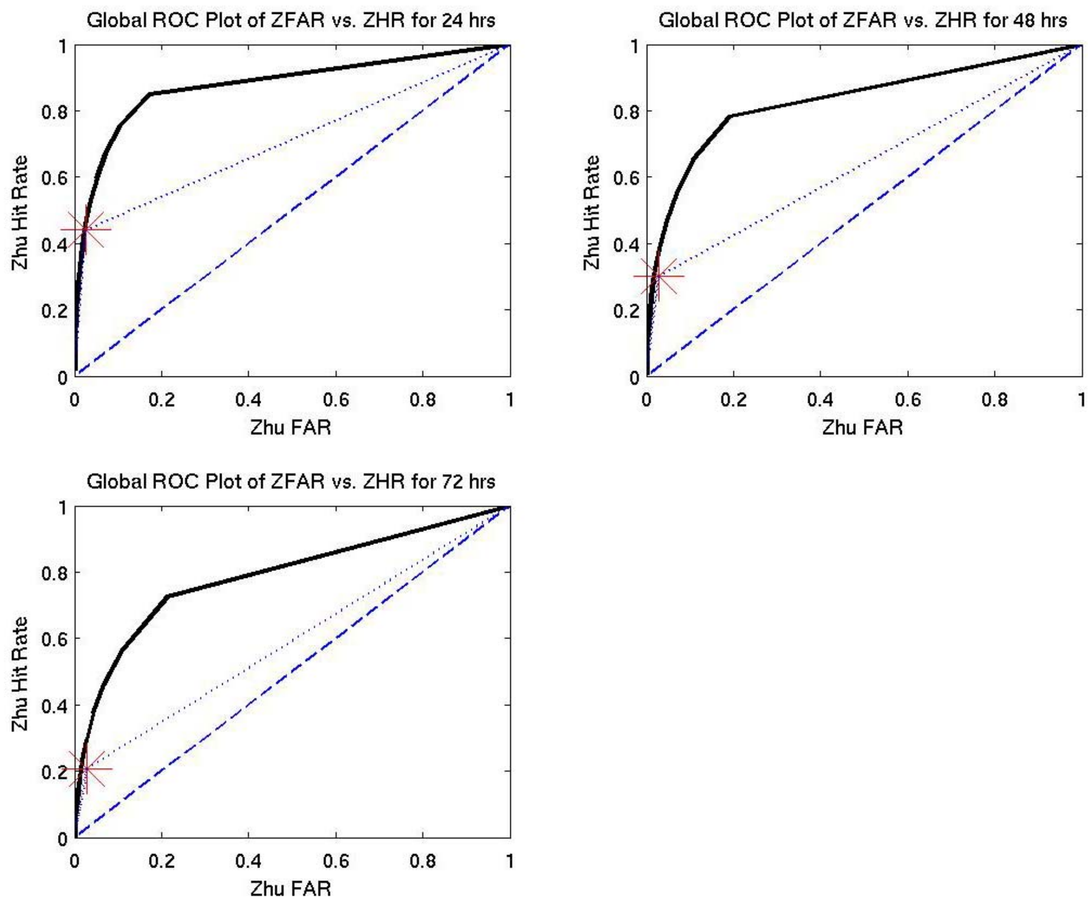
a) b)



c)

Figure 23. Economic value vs. C-L ratio plots for Case 3 of the probabilistic model (System A -blue solid line) and the deterministic model (System B – red dashed line) for (a) 24 hour, (b) 48 hour forecasts, and (c) 72 hour forecasts.

a) b)



c)

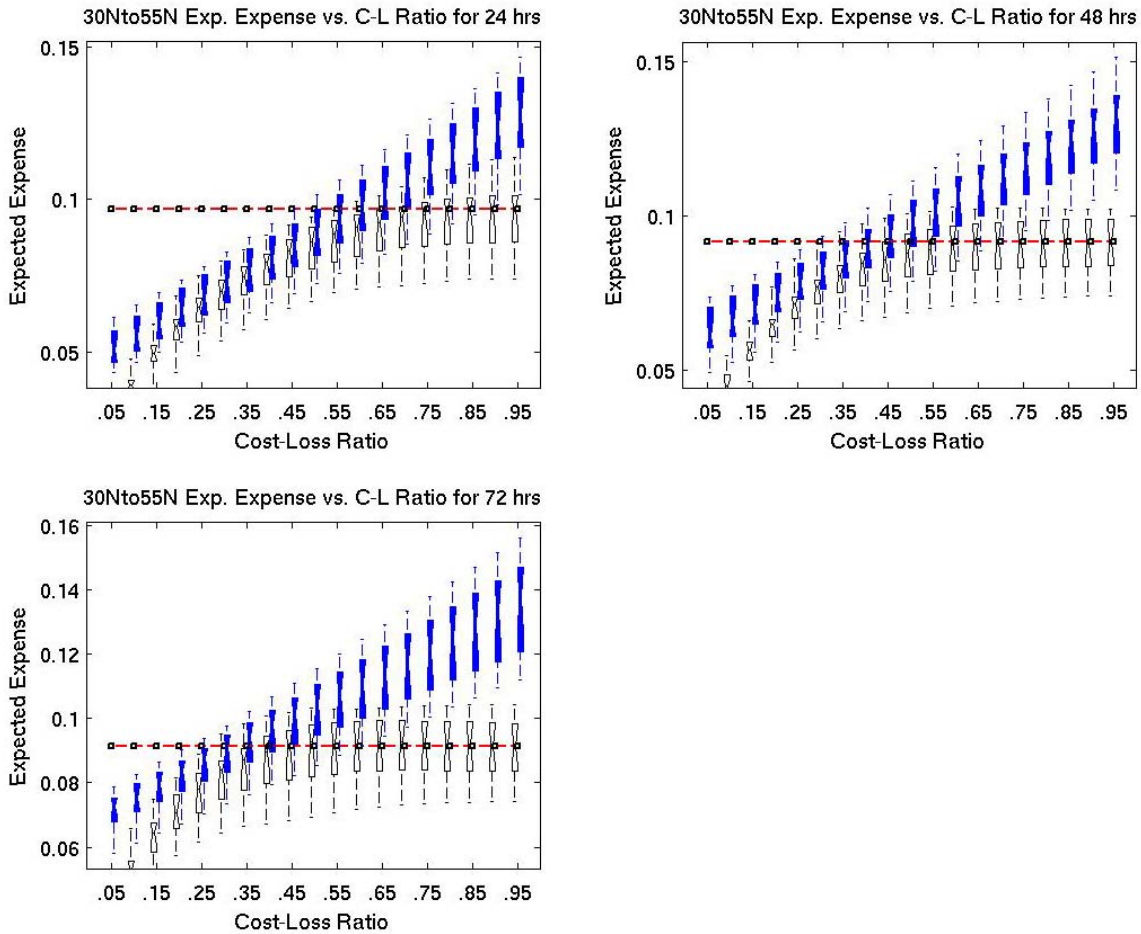
Figure 24. The ROC diagrams for Case 3, at (a) 24 hour, (b) 48 hour, and (c) 72 hour forecasts. The probabilistic system A is defined by the solid black line and the deterministic system is represented by the red star.

5. Case 4

Case Study 4 was limited to 30 N to 55 N to focus on the influences from the polar-front jet stream. At 24 hours, the difference in the mean of expected expenses for System A (probabilistic) and System B (deterministic) were not statistically significant for C-L ratios of 0.25 to 0.55 (Appendix C). At 48 hours, the null hypothesis cannot be rejected for C-L ratios 0.25 to 0.4. At 72 hours, the null hypothesis cannot be rejected for a C-L ratio of 0.25. Figure 25 illustrates the expected expense for each forecast system at different C-L ratios for Case 4.

At 24 hours (Figure 26a) there is little relative economic value of System A over system B for C-L ratios 0.25-0.55. However, for many forecast users with high and low C-L ratios, System A has significantly more economic value than System B. By 48 hours (Figure 26b), the value of System A begins to increase over the value of System B. By 72 hours (Figure 26c), System A has more relative economic value than System B for most forecast users based on C-L ratios. Again, Figure 27 demonstrates the inherent increase in utility that probabilistic forecasts have over deterministic forecasts. Probabilistic forecast systems have more utility because more forecast users can take advantage of the multiple decision levels.

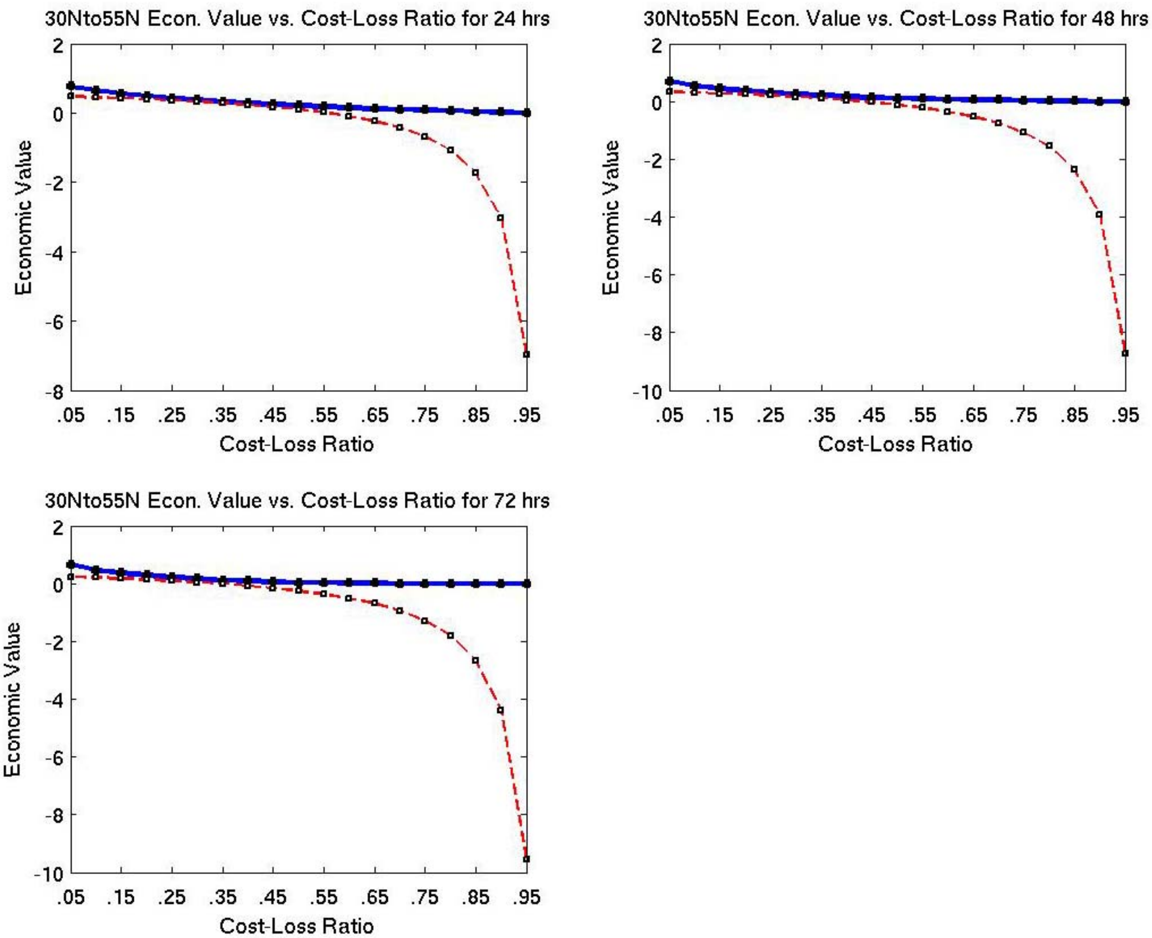
a) b)



c)

Figure 25. The expected expense value plots for Case 4 at (a) 24 hour, (b) 48 hours, and (c) 72 hours. The probabilistic model (System A) is defined by the unshaded box plots, the deterministic model (System B) is defined by the blue shaded box plots, and the median expected expense of not protecting is defined by the red dashed line.

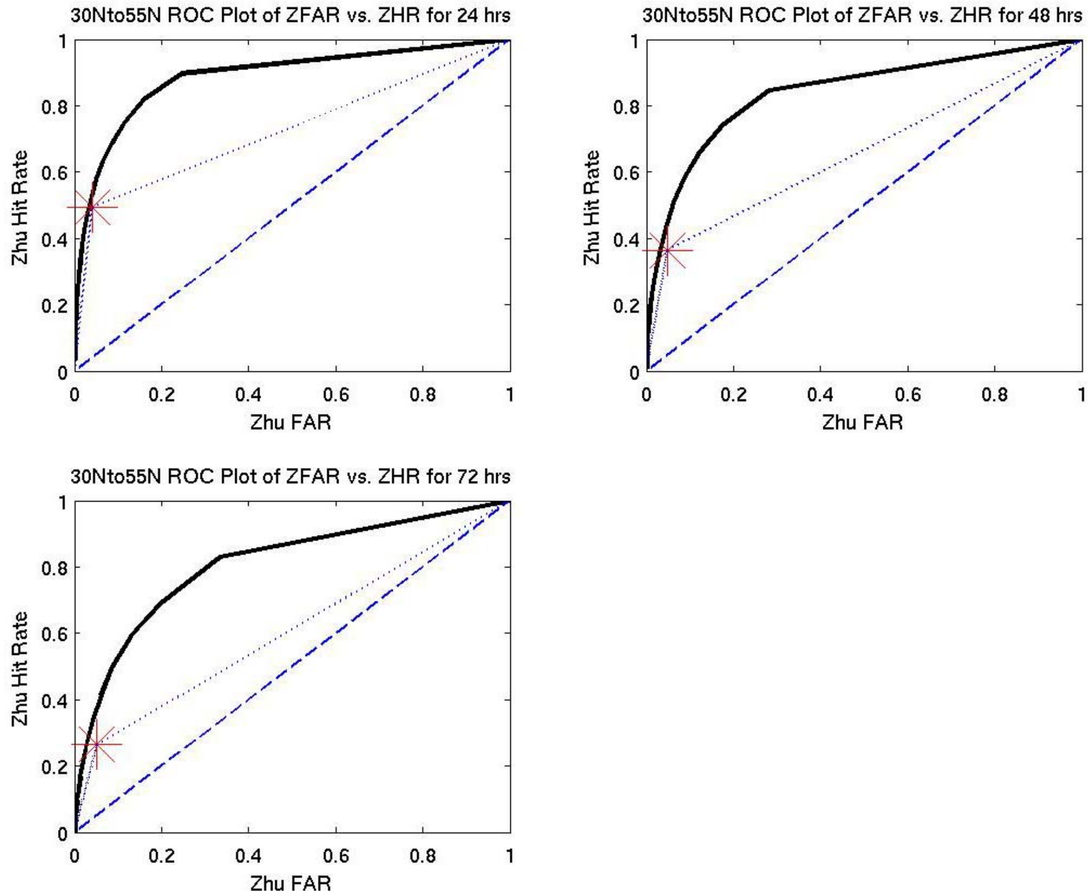
a) b)



c)

Figure 26. Economic value vs. C-L ratio plots for Case 4 of the probabilistic model (System A -blue solid line) and the deterministic model (System B – red dashed line) for (a) 24 hour, (b) 48 hour, and (c) 72 hour forecasts.

a) b)



c)

Figure 27. The ROC diagrams for Case 4, at (a) 24 hour, (b) 48 hour, and (c) 72 hour forecasts. The probabilistic system A is defined by the solid black line and the deterministic system is represented by the red star.

6. Summary

As expected, the results generated for Objective 2 are generally consistent with the results in Zhu et al. (2002). The analysis demonstrates that more rational forecast users classified by C-L ratio will benefit from forecast System A (probabilistic) than forecast System B (deterministic) for 24 to 72 hour forecasts. The relative benefit increases with increased forecast times. At 24 hours, System A could not be shown to have more relative economic value than System B for forecast users with C-L ratios from approximately 0.15 to 0.6. However, System A had more relative economic value for

users with C-L ratios less than 0.15 and greater than 0.6. Overall, more forecast users (more C-L ratios) were able to benefit from 48 hour and 72 hour forecasts. Cases 1, 3, and 4 demonstrated an increasing relative economic value of System A over system B with increased forecast time. However, with Case 2 increased relative economic value with forecast time could not be determined with statistical certainty.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. INTEGRATING PROBABILISTIC TURBULENCE FORECAST INFORMATION INTO THE AIR FORCE DECISION-MAKING PROCESS

A. OVERVIEW

Background research and the results from this thesis demonstrate that overall probabilistic forecast systems have more relative economic value than deterministic forecast systems for some forecast users with respect to their C-L ratio. Demonstrations were conducted with the assumption that forecast users were always willing to act on the basis of expected monetary value. The results suggest that there is economic advantage for the forecast user to employ reliable probabilistic forecast information over deterministic forecast information. Some users, based on C-L ratio, realize little or no gain from using probabilistic forecast information over deterministic forecast information. However, those forecast users are limited to a small range of C-L ratios and these results depend on forecast length.

Generally, the forecast user should understand their C-L ratio to react appropriately to a probabilistic forecast. However, often times a forecast user's C-L ratio is not immediately apparent without major effort. In addition, some forecast users' C-L ratios are influenced by priority, risk, and other intangible variables, which cannot be easily quantified. For many civilian applications, a strictly economic approach may be used, but for military applications there are often other non-monetary factors too difficult to characterize in terms of money, which influence the forecast user's ability to accept meteorological risks. A conceptual method, that considers operational risk management and mission priority, is proposed as an alternative to quantitatively defining C and L.

Scenarios built around the Air Force function of air refueling are analyzed. Air refueling is a critical operational function accomplished by the U.S. Air Force to maintain Air and Space Superiority and is defined as "…the in-flight transfer of fuel between tanker and receiver aircraft" (AFDD1, 2003). Air refueling (AR) significantly enhances the U.S. Air Force's ability to complete other missions critical to maintain national security. They include missions such as: "nuclear operations support, global strike, airbridge support, aircraft deployment, theater support, and special operations support"

(AFDD1, 2003). The scenarios are designed to demonstrate the use of probabilistic turbulence forecasts as tools for reducing risk while conducting the important Air Force Air and Space Function air refueling.

B. CHARACTERIZING THE FORECAST USER

1. Defining C and L by Traditional Means

Scenarios are defined relative to a forecast user who is a mission commander of a KC-135 tanker, which is capable of air refueling. Characterizing the forecast user is difficult and can become complex quickly.

Recall, C is the cost the forecast user incurs as a result of protecting due to a forecast event (in this case, the forecast of moderate or greater aircraft-scale turbulence). The loss, L, is associated with not protecting when a weather event occurs. For idealized situations, C and L are defined and fixed for a forecast user. Unfortunately, that is not the case for the AR community. An understanding of air refueling missions is essential to appropriately characterize the forecast user based on the C-L ratio. Each air refueling mission has its own unique expense characteristics, so C and L are mission dependent.

An intelligent pilot will avoid a total loss by altering their mission profile. Typically, if the forecast user encounters unforecast moderate or severe turbulence in an AR track, anecdotal evidence suggests they will alter the mission profile by changing altitudes, offsetting geographically, extending their track, or canceling the mission (Lt. Col. B. Davis, 2006, personal communication). The mission alterations are listed in decreasing order of likelihood. Occasionally, aircraft scale turbulence does contribute to aircraft mishaps (damage to aircraft or people), which would contribute to a loss, L. One might think that L should include the total loss of the aircraft or aircraft mishap expenses. However, a large majority of the time, a large loss is unlikely because the forecast user will alter the mission before a large loss occurs.

Instead of using the aircraft and mishaps in the total loss, the forecast user probably use the expenses associated with altering the mission as the loss value. Mission expense is a function of time, M(t), which includes aircraft maintenance, fuel, personnel costs, etc. In such cases, the expenses incurred as a result of the mission alteration will

be a function of time, as well. The cost, C, would be defined as the expenses associated with the forecast user performing the altered mission minus the expenses of the original mission had the forecast user performed the original mission. For example, if a forecast user had planned to fly to an altitude of 39k feet to conduct an AR mission, but unforecast moderate to severe turbulence prohibited the forecast user from using the area, they may choose to alter their mission profile. Perhaps, they took the time to find an altitude with sufficiently calm air. Suppose they found sufficiently calm air at 32k feet. There would be an increase in flight time, which would increase mission related expenses. An altered mission, $M_{a,}$ would increase the total mission time, which would be greater than the original mission expense M_o. Instead, had the forecast user chosen to fly to 32k feet initially, this preferred mission would have been less expensive because the mission time would be less. The preferred mission is denoted as M_p. It becomes apparent that mission expenses are dynamic and are nearly incalculable for realistic air refueling scenarios. In a real situation, an AR forecast user will not be able to perform detailed analysis to define their C-L ratio in terms of their mission expenses. Additionally, AR forecast users do not make decisions on preserving assets alone. Instead they make decision by balancing mission priority with operational risk management.

2. Defining C and L through Operational Risk Management and Mission Priority

ORM has become an institutionalized way of thinking in the U.S. Air Force (AFPAM 90-902). Most flying communities are required to assess their operations risk through the use of worksheets or checklists. Appendix D contains figures of an example worksheet used by the 101st Air Refueling Wing (Maine Air National Guard). While the exact methods by which aircrew throughout the U.S. Air Force assess risk differ, all methods combine for an overall risk level for the mission. The overall risk level includes human factors, such as aircrew stress, fatigue, experience, etc. In addition, mission complexity, tactics, weather, and mission priority contribute to the overall risk level. Some risk factors cannot be mitigated or lowered due to their characteristics. For example, typically the mission priority is dictated by higher authority and is unable to be changed. Also, only certain aircrew may be available for a mission due to uncontrollable reasons.

The example ORM worksheet allows the aircrew to total their overall risk level into points. The score ranges from 0 to 215. Scores from 0 to 49 are deemed low risk, scores of 50 to 80 are cautioned, and a risk score of 81 or more is considered high risk. Individual risk factors are weighted by their potential impact to the mission. For example, moderate turbulence or moderate icing in-route contributes 15 points to the overall risk level. Fortunately, unlike some risk factors some weather mission impacts can possibly be mitigated through probabilistic forecasts. For example, if an overall mission risk level is 81 the worksheet requires the aircrew to attempt to reduce their risk level. Risk mitigation could be accomplished by changing altitudes, flight path, etc. to avoid areas where weather impacts have a higher forecast probability or the probabilistic forecast may indicate that the certainty of the forecast event is low enough to lower the overall risk score. With a deterministic forecast, the forecast user is required to evaluate risk points based on the worst case scenario. Alternatively, if a forecast user is highly sensitive to a weather impact they may want to avoid areas where there is even a small chance of the weather event occurring. Such weather event certainties are not available through deterministic forecasts. Generally, aircrew do not cancel a mission because of in-route weather (Lt. Col. B. Davis, 2006, personal communication), especially if the mission is of high enough priority. The second page of the ORM worksheet (Appendix D) lists mission priority levels. The aircrew will attempt to mitigate the weather risk by performing some action to avoid a loss. Probabilistic forecasts can help them mitigate the weather risk.

In the ORM framework, the C-L ratio should be thought of in non-monetary units. The C and L should be thought of as functions of money, risk and mission priority. Mission priority is valued higher than mission risk. In Figure 28, C can be though of as an average cost of protecting that is held constant. Realistically C is mission dependent. In the figure, L adjusts based on mission priority and overall mission risk. Figure 28 is a conceptual look at how a C-L ratio in an ORM framework works. Since the C-L ratio for military forecast users is dependent on the qualitative values of risk and mission priority, it is difficult to know exactly where each scenario lies on the C-L ratio scale. However, it is possible to know approximately where each scenario is with respect to the other scenarios. Figure 28 assumes that the aircrew will avoid a loss if unforecast turbulence is

encountered. Unfortunately, sometimes the aircraft mishaps do occur (not accounted for in the Figure 28 or the scenarios). Four scenarios were devised to demonstrate how one would effectively mitigate risk with a probabilistic forecast of aircraft-scale turbulence.

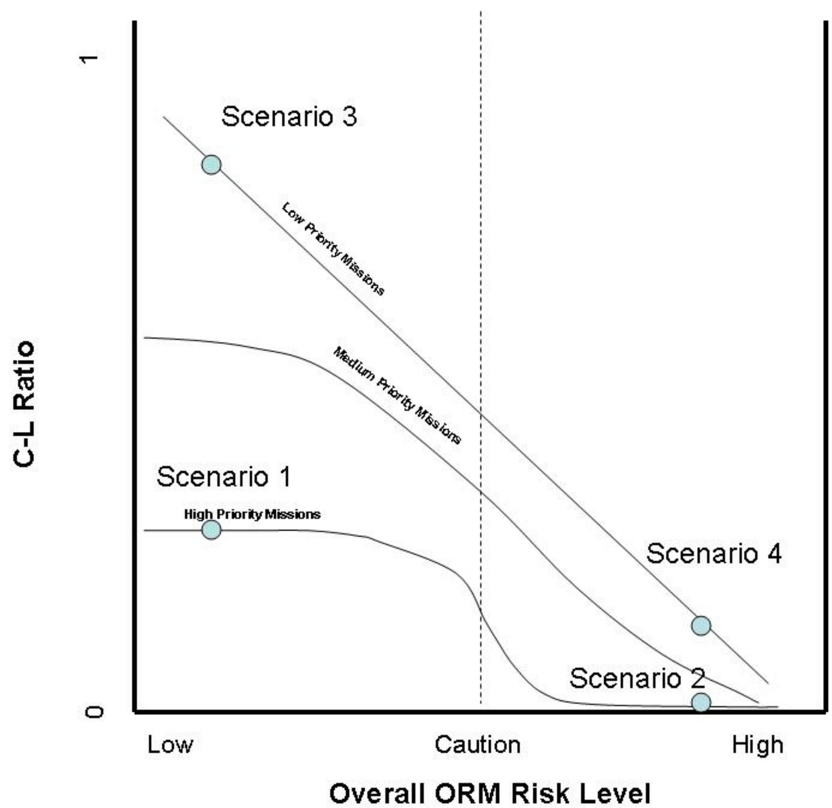


Figure 28. C-L ratio vs. Overall ORM Risk Level low, medium and high priority missions (Scenario 1 – High Priority/Low Overall Risk, Scenario 2 – High Priority/High Overall Risk, Scenario 3 – Low Priority/Low Overall Risk, Scenario 4 – Low Priority/High Overall Risk).

C. RESULTS

1. Risk Mitigation

a. Scenario 1 – High Priority/Low Overall Risk

Scenario 1 is an example of a forecast user characterized with a high mission priority and a low overall risk level according to their ORM framework. Figure 29 lists possible mission characteristics for a high priority/low overall risk scenario. Figure 30 illustrates example probabilistic and deterministic forecasts and how they may

look in such a scenario. In this scenario, when the forecast user relies on a deterministic forecast they initially fly to an alternate AR track to accomplish the mission. Given the high priority of the mission, it is logical that the forecast user would want to complete the mission even if that meant a slight delay. For the traditional deterministic forecast scenario, the forecast user chose to ignore the forecast and attempted the AR. The AR was accomplished but caused a mission delay and incurred additional mission expenses. For the future probabilistic forecast the forecast user chose to go to an AR track with less probability of moderate to severe turbulence. Unfortunately, the forecast user encountered mission impacting turbulence at this alternate location. Additional expenses were incurred as a result of needing to find calm air. This scenario illustrates that probabilistic forecast will not eliminate false alarms and misses. The C-L ratio for Scenario 1 (see Figure 28) should be in the bottom half of the C-L ratio spectrum. A high priority mission positively contributes to the L value, thus causing C/L to decrease.

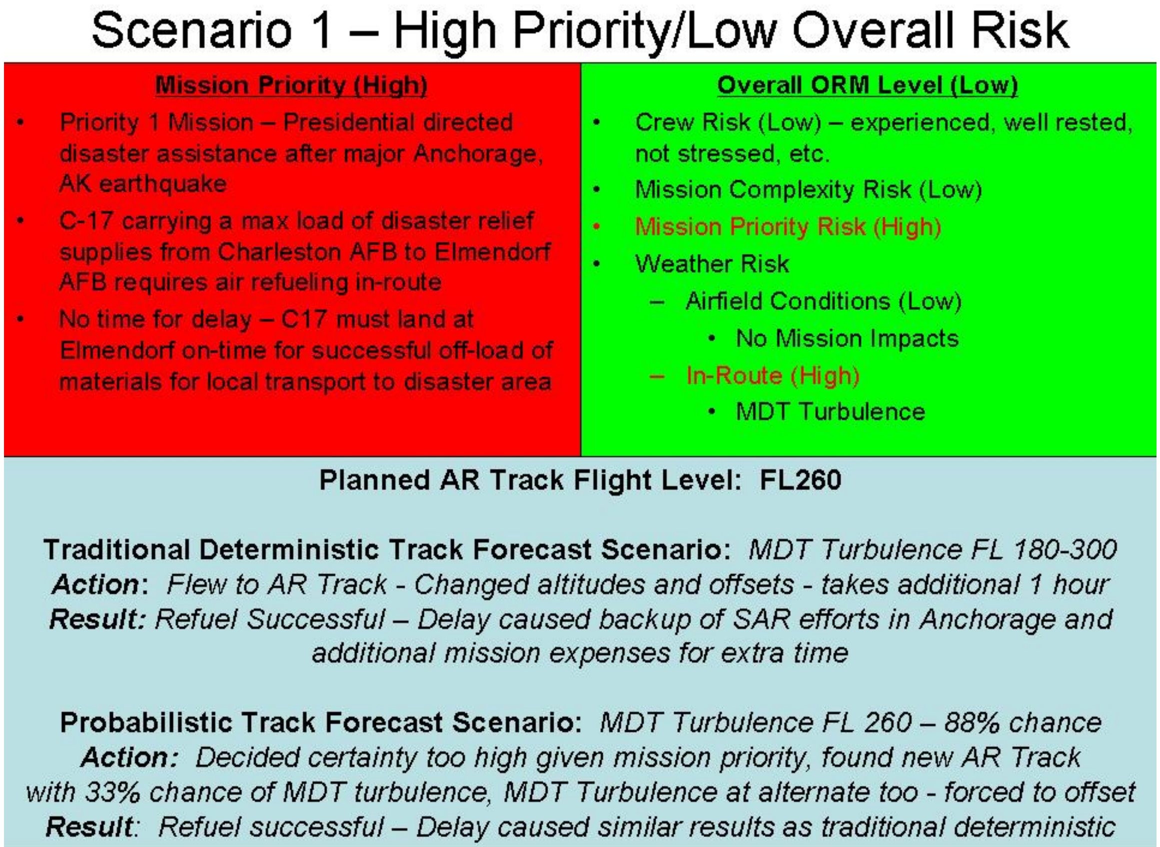


Figure 29. Scenario 1 – Description.

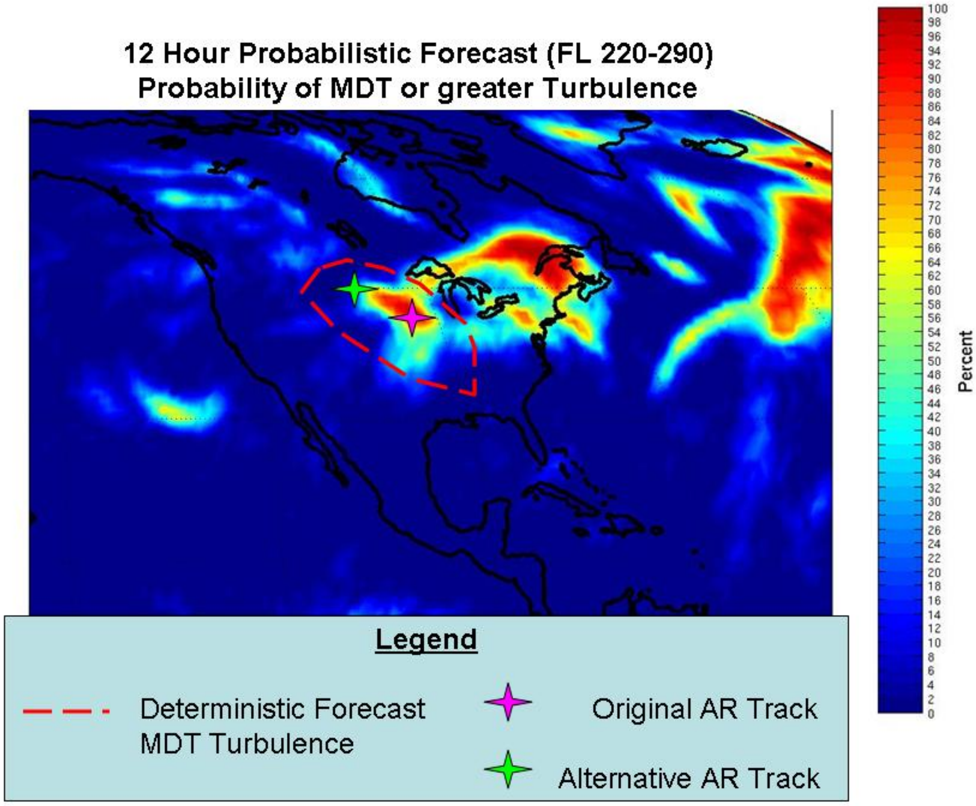


Figure 30. Scenario 1 – Probabilistic and Deterministic Forecasts.

b. Scenario 2 – High Priority/High Overall Risk

Scenario 2 is an example of a forecast user characterized with a high mission priority and a high overall risk level according to their ORM framework. Figure 31 lists possible mission characteristics for a high priority and high risk mission. Figure 32 illustrates example deterministic and probabilistic forecasts for a scenario and where the AR tracks might be located for the given scenario. In the traditional deterministic forecast scenario, the forecast user decides to use the original AR track because moderate or greater turbulence is not forecast for that track. The traditional deterministic forecast lacks important uncertainty information. For example, the human forecaster who made the turbulence forecast might have thought that there was a slight chance for turbulence at the original AR track location, but did not communicate that to the CWT forecaster who prepared the pilot's weather brief. In turn, the CWT forecaster never relayed any

certainty information to the pilot. In the future probabilistic scenario, all parties involved would have been put on alert that the model indicates a reasonable chance for moderate or greater turbulence. In the future probabilistic scenario, the pilot was able to choose a location that better suited his tolerance for risk given his mission priority. Of all possible scenarios, a high mission priority and high overall risk scenario would likely result in the lowest C-L ratio. This occurs because high mission priority and high risk contribute positively to L, which decreases C-L. Since Scenario 2 forecast users have such a low C-L, they should react to forecasts when there is even a slight chance of an event occurring (Figure 28).

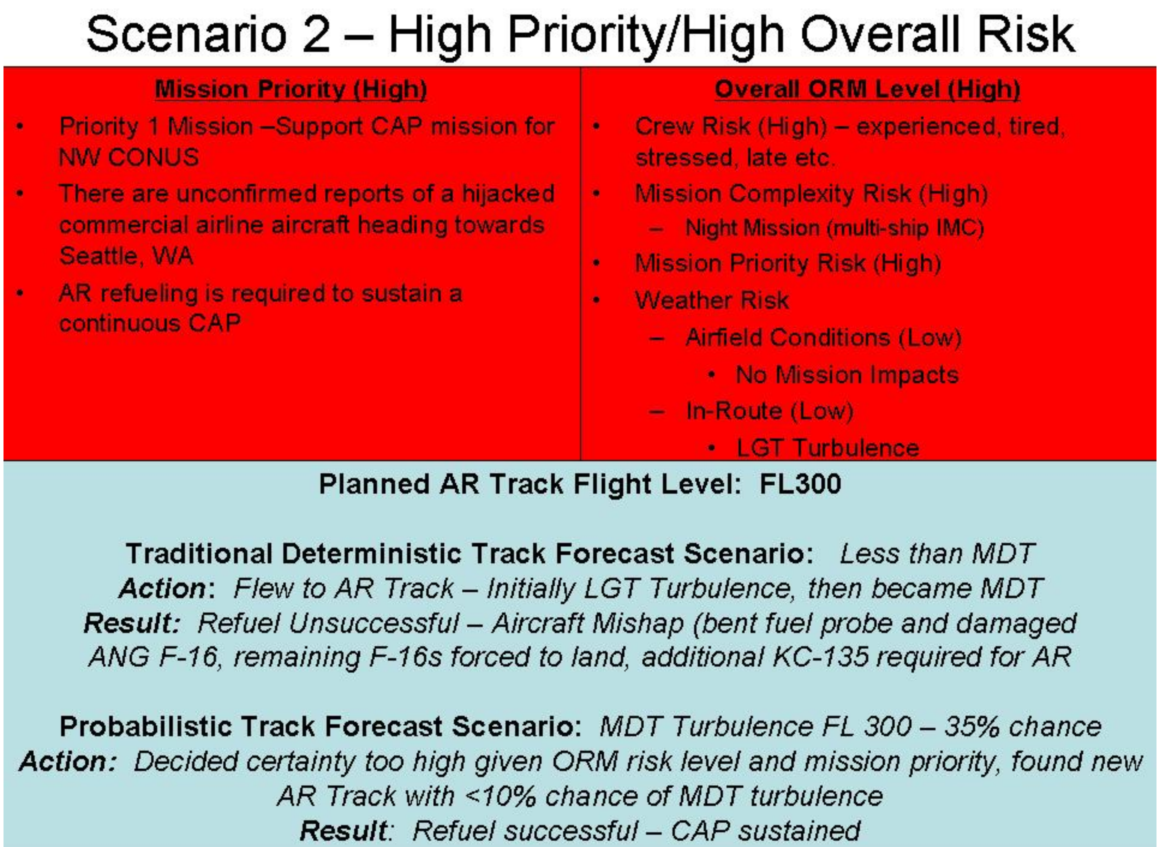


Figure 31. Scenario 2 – Description.

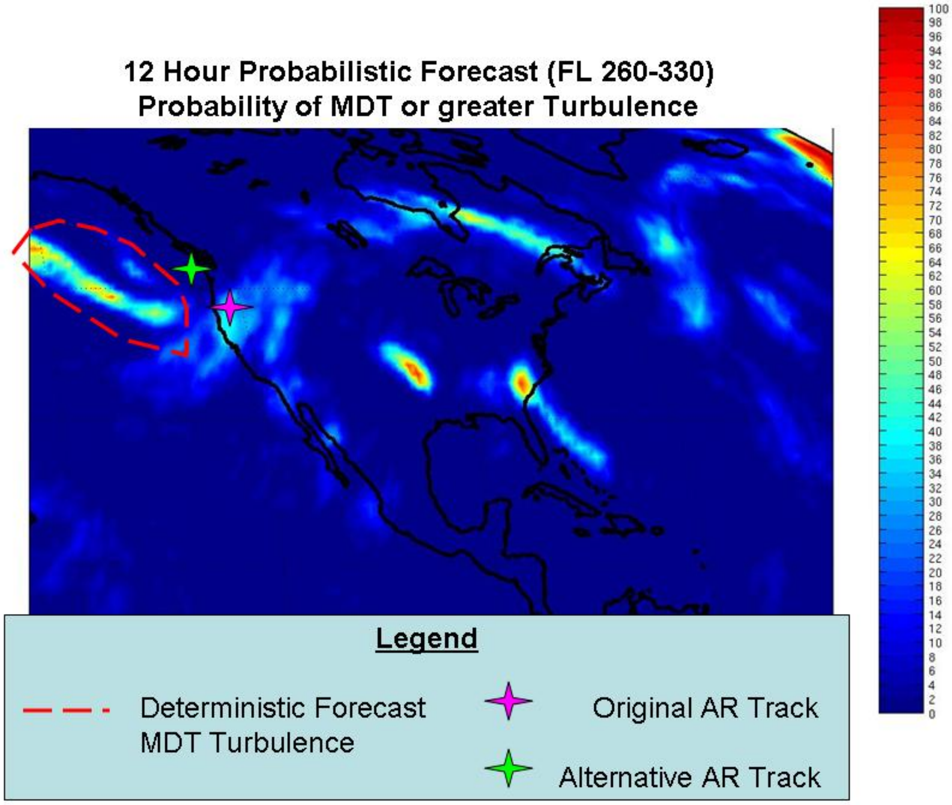


Figure 32. Scenario 2 – Probabilistic and Deterministic Forecasts.

c. Scenario 3 – Low Priority/Low Overall Risk

Scenario 3 is an example of a forecast user characterized with a low mission priority and a low overall risk level according to their ORM framework. Figure 33 lists possible mission characteristics for a low priority and low risk mission. Figure 34 illustrates example deterministic and probabilistic forecasts for a scenario and where the AR tracks might be located for the given scenario. In the traditional deterministic forecast scenario, the forecast user chose to fly to an alternate AR track instead of attempting the original AR track where MDT turbulence was forecast. The decision incurred additional mission related expenses. With the future probabilistic forecast scenario the forecast user chose to try the original AR track anyhow, because the low mission priority and low risk meant that they would not lose much by trying the original AR track. It happened that only light turbulence was encountered at the original AR

track. No additional expenses were incurred. Scenario 3 users would have a high C-L ratio compared to all other scenarios, meaning that they should choose to react to forecasts with high event certainty (Figure 28).

Scenario 3 – Low Priority/Low Overall Risk

Mission Priority (Low)	Overall ORM Level (Low)
• Priority 3 Mission –Training Mission	• Crew Risk (Low) – experienced, not stressed, not tired, etc. • Mission Complexity Risk (Low) • Mission Priority Risk (Low) • Weather Risk – Airfield Conditions (Low) • No Mission Impacts – In-Route (High) • MDT Turbulence

Planned AR Track Flight Level: FL280

Traditional Deterministic Track Forecast Scenario: *MDT Turbulence FL 250-350*
Action: *Flew to alternate AR Track*
Result: *Training Successful, longer mission than needed, extra mission planning, incur additional costs for flying to alternate AR track*

Probabilistic Track Forecast Scenario: *MDT Turbulence FL 280 – 40% chance*
Action: *Given low mission priority and risk, go to original AR track can change altitudes or offset if need too, only LGT turbulence encountered*
Result: *Training successful, no additional mission expenses*

Figure 33. Scenario 3 – Description.

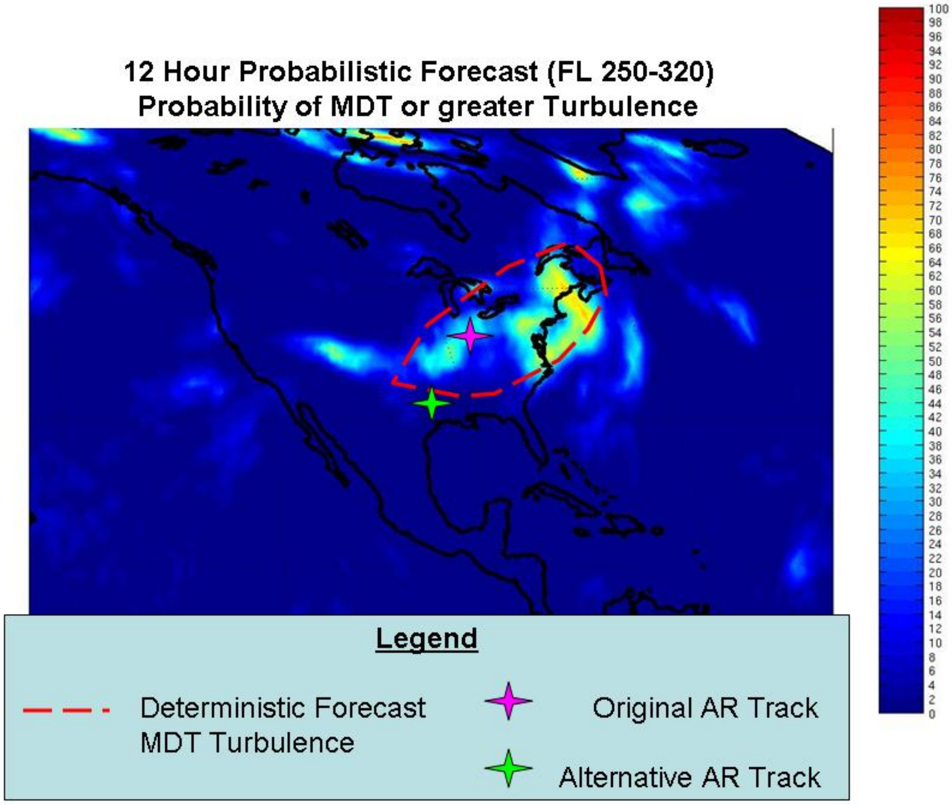


Figure 34. Scenario 3 – Probabilistic and Deterministic Forecasts.

d. Scenario 4 – Low Priority/High Overall Risk

Scenario 4 is an example of a forecast user characterized with a low mission priority and a low overall risk level according to their ORM framework. Figure 35 lists possible mission characteristics for a low priority and high risk mission. Figure 36 illustrates example deterministic and probabilistic forecasts for a scenario and where the AR tracks might be located for the given scenario. In the traditional deterministic forecast scenario, the forecast user chose to fly to the alternate AR track because moderate turbulence was forecast. The future probabilistic forecast scenario forecast user also chose to fly to the alternate AR track, but was able to make the decision based on forecast certainty and their particular mission priority and risk level. This scenario demonstrates that in some cases, the results from using a deterministic forecast and probabilistic forecast may be the same. The C-L ratio for Scenario 4 forecast users will

be higher than Scenario 2 forecast users, because mission priority significantly contributes to the L for Scenario 2 forecast users (Figure 28).

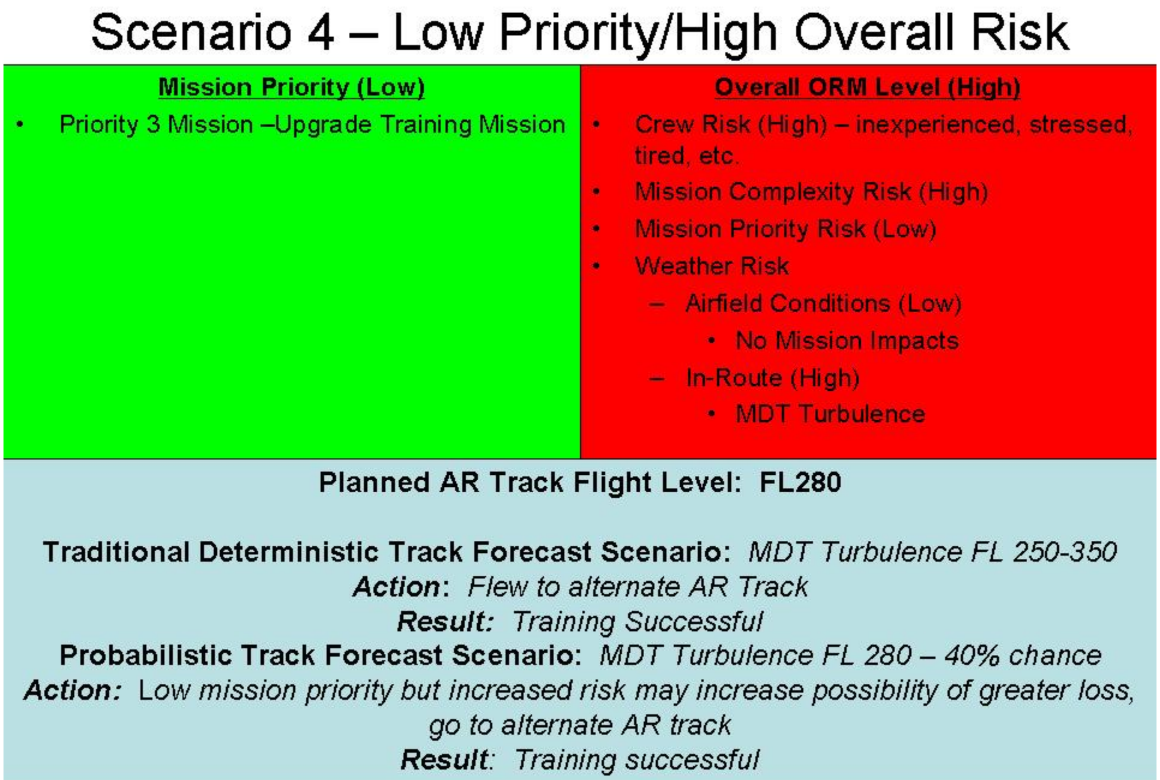


Figure 35. Scenario 4 – Description.

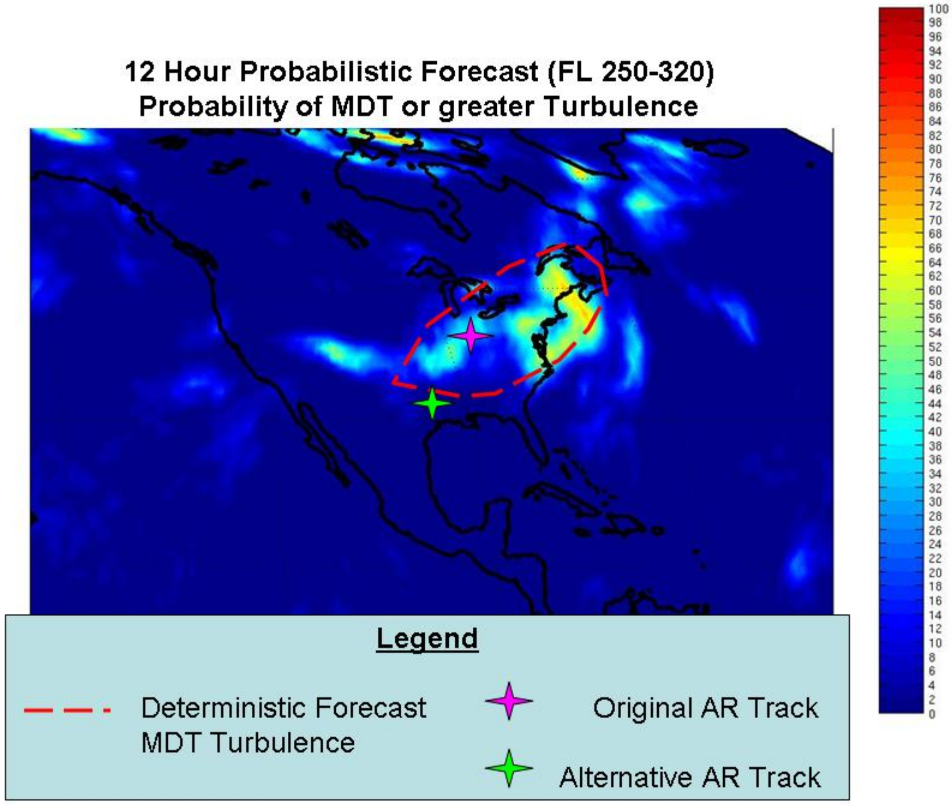


Figure 36. Scenario 4 – Probabilistic and Deterministic Forecasts.

2. Summary

The previous scenarios demonstrate a few possible outcomes from using both a probabilistic forecast system and a deterministic forecast system. They do not represent all possible situations or outcomes and should only be taken as examples. The scenarios demonstrate realistic potential benefits to using stochastic forecasts over deterministic forecasts. For many military forecast users, it is difficult to quantify their C-L ratio. By applying Figure 28 to the scenarios, one is able to see a qualitative relationship between C-L ratio, mission priority and mission risk. C and L do not need to reflect only monetary values and may reflect other non-quantifiable values such as mission priority and mission risk. Stochastic forecasts do not eliminate false alarms or misses, but they have been demonstrated to have more utility than deterministic forecasts (i.e., Objective 2).

VII. CONCLUSION

A. FINAL REMARKS

The three objectives of this thesis were to: (1) create an ensemble-based turbulence forecast system capable producing forecast probability for air turbulence that impacts flight operations for this thesis, (2), to demonstrate the advantages of providing forecasts based on probability of occurrence over traditional deterministic forecasts (3) and to integrate probabilistic turbulence forecast information into the Air Force decision-making process.

The creation of the ETFS required an understanding of several scientific disciplines, to include: meteorology (aircraft-scale turbulence), statistics and probability, and numerical weather prediction. A well-designed ensemble prediction system should account for initial condition uncertainty and model error. In this thesis GFS ensemble members were chosen for their availability and better methods for accounting initial condition uncertainty may be available for an operational ETFS. Additionally, limited ensemble members from the GFS (5 positive and 5 negative perturbations), required a lagged-average forecasting approach to create more than 10 ensemble members. A version of the Ellrod Turbulence Index was used to post-process atmospheric variables from the 40 ensemble members to generate turbulence forecasts for each grid point over the globe. The uniform ranks method was used to calculate forecast probability. A better method, called weighted ranks (calibrated ensemble members), is suggested for future ETFS work. The ETFS designed in this thesis was sufficient for the purposes of exploring the second and third objectives.

Zhu et al. (2002) and others have demonstrated how ensemble-based probabilistic forecasts have greater utility than deterministic forecasts. Thesis results support their assertions and clearly demonstrate the overall advantage of using ensemble-based probabilistic forecasts versus deterministic forecasts, in the long run. An ETFS was created to examine their assertions. The probabilistic forecast system (System A) provided more value than the deterministic forecast system (System B) to forecast users with high and low C-L ratios. Forecast users with a low middle-range C-L ratio did not

benefit more from System A than System B in short-range forecasts. Those users do begin to see some advantages at later forecast times.

Finally, scenarios were created to demonstrate how the integration of stochastic forecast guidance into the Air Force decision-making process might be accomplished. Statisticians and decision-theorists know that decision-makers who act on the basis of an expected monetary value for the long-run will have the largest savings at the end. Military planners and decision-makers may not explicitly state this as an assumption in their decision-making. However, military planners and decision-makers do implicitly make decisions in this manner. Unfortunately, military planners and decision-makers do not necessarily speak the same language. For example, military decision-makers do not necessarily keep account of their C and L in directly quantifiable or monetary terms. Instead, the U.S. Air Force prefers to operate with a culture defined by Operational Risk Management and mission priority. In this thesis, the C-L ratio used by statisticians and meteorologists was related to mission priority and operational risk level (Figure 28). A qualitative understanding of the relationship between C-L ratio and mission priority and risk is essential for the DoD to integrate and apply ensemble-based probabilistic forecasts into its operations.

B. SUGGESTIONS FOR FUTURE RESEARCH

A more robust and reliable ETFS should be fielded for operational testing. This can be accomplished by employing more sophisticated techniques to generate ensemble members, to include: using higher resolution model output, using a varied-model technique (Eckel and Mass 2005), calibrating ensemble members, and creating a strong verification system for turbulence that takes advantage of new turbulence observation techniques. The varied-model technique implies using multiple diagnostic methods for diagnosing turbulence and weighting those ensemble members according to how well they perform. Additionally, a working group of DoD meteorologists, aviators and operations analysts needs to be established to appropriately address integrating probabilistic forecasts into DoD operations. A broad view of the decision process should be considered when integrating stochastic or deterministic weather information into operations.

APPENDIX A: LAGGED-AVERAGE FORECASTING TABLE

The tables in this appendix detail how the lagged average forecasting was handled for the ETFS. Each ensemble run will be considered to have a run time of 18Z. Ensemble members based on 00Z, 06Z, and 12Z are used to increase the number of ensemble members in the ETFS. For example, for a 06-h forecast from an 18Z ensemble run, the forecast time for the old runs will be 24 hr, 18 hr, and 12 hr forecast for 00Z, 06Z, and 12Z runs respectively. All model runs are on the same day.

	Run Time (Z)	Forecast Hour
6 Hour Ensemble Forecast	18	06
	12	12
	06	18
	00	24
12 Hour Ensemble Forecast	Run Time (Z)	Forecast Hour
	18	12
	12	18
	06	24
	00	30
18 Hour Ensemble Forecast	Run Time (Z)	Forecast Hour
	18	18
	12	24
	06	30
	00	36

24 Hour Ensemble Forecast	Run Time (Z)	Forecast Hour
	18	24
	12	30
	06	36
	00	42
30 Hour Ensemble Forecast	Run Time (Z)	Forecast Hour
	18	30
	12	36
	06	42
	00	48
36 Hour Ensemble Forecast	Run Time (Z)	Forecast Hour
	18	36
	12	42
	06	48
	00	54
48 Hour Ensemble Forecast	Run Time (Z)	Forecast Hour
	18	48
	12	54
	06	60
	00	66

	Run Time (Z)	Forecast Hour
54 Hour Ensemble Forecast	18	54
	12	60
	06	66
	00	72
72 Hour Ensemble Forecast	Run Time (Z)	Forecast Hour
	18	72
	12	78
	06	84
	00	90

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B: ELLROD INDEX THRESHOLD CALIBRATION

Appendix B contains contingency tables and other data used in analysis for Objective 1 to generate results for determining which Ellrod turbulence diagnostic threshold to use for Objective 2 analysis. The POD, TS, HR, FAR, Bias, HSS, Zhu HR, and Zhu FAR used in this section refer to the definitions outlined in Table 3.

```
TI>1 Results

              Observed Yes   Observed No
Forecast Yes   32            394
Forecast No    19            313

n = 758

POD = 0.627451    FAR = 0.924883
TS = 0.071910     Bias = 8.352941
HR = 0.455145     HSS = 0.015906


TI>1_5 Results

              Observed Yes   Observed No
Forecast Yes   30            302
Forecast No    21            405

n = 758

POD = 0.588235    FAR = 0.909639
TS = 0.084986     Bias = 6.509804
HR = 0.573879     HSS = 0.151849


TI>2 Results

              Observed Yes   Observed No
Forecast Yes   27            222
Forecast No    24            485

n = 758

POD = 0.529412    FAR = 0.891566
TS = 0.098901     Bias = 4.882353
HR = 0.675462     HSS = 0.076900
```

TI>3 Results

	Observed Yes	Observed No
Forecast Yes	21	127
Forecast No	30	580

n = 758

POD = 0.411765   FAR = 0.858108
TS = 0.117978   Bias = 2.901961
HR = 0.792876   HSS = 0.123319

TI>4 Results

	Observed Yes	Observed No
Forecast Yes	16	82
Forecast No	35	625

n = 758

POD = 0.313725   FAR = 0.836735
TS = 0.120301   Bias = 1.921569
HR = 0.845646   HSS = 0.138519

TI>4_5 Results

	Observed Yes	Observed No
Forecast Yes	16	70
Forecast No	35	637

n = 758

POD = 0.313725   FAR = 0.813953
TS = 0.132231   Bias = 1.686275
HR = 0.861478   HSS = 0.151849

TI>5 Results

	Observed Yes	Observed No
Forecast Yes	15	57
Forecast No	36	650

n = 758

POD = 0.294118   FAR = 0.791667
TS = 0.138889   Bias = 1.411765
HR = 0.877309   HSS = 0.179253

TI>5_5 Results

	Observed Yes	Observed No
Forecast Yes	11	44
Forecast No	40	663

n = 758

POD = 0.215686   FAR = 0.800000
TS = 0.115789   Bias = 1.078431
HR = 0.889182   HSS = 0.151849

TI>6 Results

	Observed Yes	Observed No
Forecast Yes	10	34
Forecast No	41	673

n = 758

POD = 0.196078   FAR = 0.772727
TS = 0.117647   Bias = 0.862745
HR = 0.901055   HSS = 0.158052

TI>7 Results

	Observed Yes	Observed No
Forecast Yes	7	22
Forecast No	44	685

n = 758

POD = 0.137255   FAR = 0.758621
TS = 0.095890   Bias = 0.568627
HR = 0.912929   HSS = 0.132693

TI>8 Results

	Observed Yes	Observed No
Forecast Yes	7	16
Forecast No	44	691

n = 758

POD = 0.137255   FAR = 0.695652
TS = 0.104478   Bias = 0.450980
HR = 0.920844   HSS = 0.153797

```
TI>9 Results

                Observed Yes   Observed No
Forecast Yes      6             10
Forecast No      45            697

n = 758

POD = 0.117647    FAR = 0.625000
TS = 0.098361     Bias = 0.313725
HR = 0.927441     HSS = 0.151849
```

TI >	POD	FAR	TS	Bias	Wilks HR	HSS	Zhu HR	Zhu FAR
1	0.627451	0.924888	0.07191	8.352941	0.455145	0.015906	0.627	0.557
1.5	0.588235	0.909639	0.084986	6.509804	0.573879	0.151849	0.5882	0.427
2	0.529412	0.891566	0.098901	4.882353	0.675462	0.0769	0.529	0.314
3	0.411765	0.858108	0.117978	2.901961	0.792876	0.123319	0.4117	0.1796
4	0.313725	0.836735	0.120301	1.921569	0.845646	0.138519	0.3137	0.11598
4.5	0.313725	0.813953	0.132231	1.686275	0.861478	0.151185	0.3137	0.099
5	0.294118	0.791667	0.138889	1.411765	0.877309	0.179253	0.294	0.0806
5.5	0.215686	0.8	0.115789	1.078431	0.889182	0.151849	0.217	0.0622
6	0.196078	0.772727	0.117647	0.862745	0.901055	0.158052	0.1961	0.048
7	0.137255	0.758621	0.09589	0.568627	0.912929	0.132693	0.13725	0.0311
8	0.137255	0.695652	0.104478	0.45098	0.920844	0.153797	0.13725	0.0226
9	0.117647	0.625	0.098361	0.313725	0.927441	0.151849	0.1176	0.01414

TI >	ROCA	ROCS
1	0.535	0.07
1.5	0.5806	0.1612
2	0.6075	0.215
3	0.61605	0.2321
4	0.59886	0.19772
4.5	0.60735	0.2147
5	0.6067	0.2134
5.5	0.5774	0.1548
6	0.57405	0.1481
7	0.553075	0.10615
8	0.557325	0.11465
9	0.55173	0.10346

APPENDIX C: HYPOTHESIS TESTING RESULTS

Detailed hypothesis testing results for Objective 2.

CASE 1

```
September Hypothesis Testing Results for Global
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.05)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0210 and -0.0119
Tstat = -7.7254  df = 16.0000  sd = 0.0045
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.1)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0010
CI are -0.0143 and -0.0044
Tstat = -4.0017  df = 16.0000  sd = 0.0050
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.15)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.0690
CI are -0.0105 and 0.0004
Tstat = -1.9493  df = 16.0000  sd = 0.0055
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.2)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.4187
CI are -0.0082 and 0.0036
Tstat = -0.8301  df = 16.0000  sd = 0.0059
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.25)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 1.0000
CI are -0.0063 and 0.0063
Tstat = 0.0000  df = 16.0000  sd = 0.0063
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.3)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.9388
CI are -0.0068 and 0.0063
Tstat = -0.0780  df = 16.0000  sd = 0.0066
```

```
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.35)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.9615
CI are -0.0069 and 0.0066
Tstat = -0.0491  df = 16.0000  sd = 0.0067
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.4)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.8652
CI are -0.0074 and 0.0063
Tstat = -0.1725  df = 16.0000  sd = 0.0069
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.45)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.6554
CI are -0.0085 and 0.0055
Tstat = -0.4547  df = 16.0000  sd = 0.0070
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.5)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.4121
CI are -0.0100 and 0.0043
Tstat = -0.8422  df = 16.0000  sd = 0.0071
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.55)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.2168
CI are -0.0116 and 0.0029
Tstat = -1.2858  df = 16.0000  sd = 0.0072
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.6)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.1047
CI are -0.0133 and 0.0014
Tstat = -1.7203  df = 16.0000  sd = 0.0074
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.65)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0414
CI are -0.0152 and -0.0003
Tstat = -2.2171  df = 16.0000  sd = 0.0074
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.7)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0156
CI are -0.0172 and -0.0021
Tstat = -2.7040  df = 16.0000  sd = 0.0076
```

```
**************************************************
24 hour forecast (Cost-Loss Ratio = 0.75)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0051
CI are -0.0193 and -0.0040
Tstat = -3.2448  df = 16.0000  sd = 0.0076
**************************************************
24 hour forecast (Cost-Loss Ratio = 0.8)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0017
CI are -0.0215 and -0.0060
Tstat = -3.7760  df = 16.0000  sd = 0.0077
**************************************************
24 hour forecast (Cost-Loss Ratio = 0.85)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0005
CI are -0.0238 and -0.0081
Tstat = -4.3215  df = 16.0000  sd = 0.0078
**************************************************
24 hour forecast (Cost-Loss Ratio = 0.9)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0002
CI are -0.0262 and -0.0103
Tstat = -4.8819  df = 16.0000  sd = 0.0079
**************************************************
24 hour forecast (Cost-Loss Ratio = 0.95)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0001
CI are -0.0286 and -0.0126
Tstat = -5.4319  df = 16.0000  sd = 0.0080
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.05)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0242 and -0.0140
Tstat = -7.9199  df = 16.0000  sd = 0.0051
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.1)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0013
CI are -0.0164 and -0.0048
Tstat = -3.8927  df = 16.0000  sd = 0.0058
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.15)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.0570
CI are -0.0122 and 0.0002
Tstat = -2.0508  df = 16.0000  sd = 0.0062
```

```
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.2)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.2155
CI are -0.0105 and 0.0026
Tstat = -1.2896  df = 16.0000  sd = 0.0065
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.25)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 1.0000
CI are -0.0067 and 0.0067
Tstat = 0.0000  df = 16.0000  sd = 0.0067
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.3)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.4421
CI are -0.0095 and 0.0043
Tstat = -0.7883  df = 16.0000  sd = 0.0069
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.35)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.4144
CI are -0.0099 and 0.0043
Tstat = -0.8379  df = 16.0000  sd = 0.0071
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.4)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.3047
CI are -0.0109 and 0.0036
Tstat = -1.0603  df = 16.0000  sd = 0.0073
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.45)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.1861
CI are -0.0124 and 0.0026
Tstat = -1.3814  df = 16.0000  sd = 0.0075
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.5)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.0888
CI are -0.0140 and 0.0011
Tstat = -1.8122  df = 16.0000  sd = 0.0075
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.55)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0386
CI are -0.0156 and -0.0005
Tstat = -2.2532  df = 16.0000  sd = 0.0076
```

```
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.6)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0153
CI are -0.0174 and -0.0021
Tstat = -2.7134  df = 16.0000  sd = 0.0076
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.65)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0060
CI are -0.0193 and -0.0038
Tstat = -3.1630  df = 16.0000  sd = 0.0078
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.7)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0021
CI are -0.0213 and -0.0057
Tstat = -3.6579  df = 16.0000  sd = 0.0078
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.75)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0008
CI are -0.0233 and -0.0075
Tstat = -4.1384  df = 16.0000  sd = 0.0079
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.8)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0003
CI are -0.0254 and -0.0095
Tstat = -4.6276  df = 16.0000  sd = 0.0080
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.85)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0001
CI are -0.0277 and -0.0115
Tstat = -5.1183  df = 16.0000  sd = 0.0081
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.9)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0300 and -0.0135
Tstat = -5.5965  df = 16.0000  sd = 0.0082
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.95)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0323 and -0.0156
Tstat = -6.0539  df = 16.0000  sd = 0.0084
```

CASE 2

```
September Hypothesis Testing Results for 30Nto55N
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.05)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0254 and -0.0170
Tstat = -10.7395  df = 16.0000  sd = 0.0042
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.1)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0175 and -0.0078
Tstat = -5.5262  df = 16.0000  sd = 0.0049
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.15)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0091
CI are -0.0129 and -0.0021
Tstat = -2.9656  df = 16.0000  sd = 0.0054
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.2)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.2071
CI are -0.0101 and 0.0024
Tstat = -1.3147  df = 16.0000  sd = 0.0062
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.25)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 1.0000
CI are -0.0065 and 0.0065
Tstat = 0.0000  df = 16.0000  sd = 0.0065
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.3)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.5275
CI are -0.0093 and 0.0050
Tstat = -0.6459  df = 16.0000  sd = 0.0072
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.35)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.4925
CI are -0.0102 and 0.0051
Tstat = -0.7025  df = 16.0000  sd = 0.0077
```

```
**************************************************
24 hour forecast (Cost-Loss Ratio = 0.4)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.4090
CI are -0.0115 and 0.0049
Tstat = -0.8478  df = 16.0000  sd = 0.0082
**************************************************
24 hour forecast (Cost-Loss Ratio = 0.45)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.2646
CI are -0.0131 and 0.0039
Tstat = -1.1561  df = 16.0000  sd = 0.0085
**************************************************
24 hour forecast (Cost-Loss Ratio = 0.5)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.1440
CI are -0.0153 and 0.0024
Tstat = -1.5364  df = 16.0000  sd = 0.0089
**************************************************
24 hour forecast (Cost-Loss Ratio = 0.55)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.0691
CI are -0.0173 and 0.0007
Tstat = -1.9483  df = 16.0000  sd = 0.0090
**************************************************
24 hour forecast (Cost-Loss Ratio = 0.6)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0321
CI are -0.0197 and -0.0010
Tstat = -2.3476  df = 16.0000  sd = 0.0093
**************************************************
24 hour forecast (Cost-Loss Ratio = 0.65)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0125
CI are -0.0223 and -0.0031
Tstat = -2.8146  df = 16.0000  sd = 0.0096
**************************************************
24 hour forecast (Cost-Loss Ratio = 0.7)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0054
CI are -0.0247 and -0.0051
Tstat = -3.2135  df = 16.0000  sd = 0.0098
**************************************************
24 hour forecast (Cost-Loss Ratio = 0.75)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0022
CI are -0.0274 and -0.0072
Tstat = -3.6391  df = 16.0000  sd = 0.0101
```

```
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.8)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0009
CI are -0.0301 and -0.0095
Tstat = -4.0667  df = 16.0000  sd = 0.0103
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.85)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0004
CI are -0.0329 and -0.0118
Tstat = -4.4869  df = 16.0000  sd = 0.0106
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.9)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0001
CI are -0.0359 and -0.0144
Tstat = -4.9577  df = 16.0000  sd = 0.0108
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.95)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0001
CI are -0.0388 and -0.0170
Tstat = -5.4165  df = 16.0000  sd = 0.0109
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.05)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0340 and -0.0191
Tstat = -7.5138  df = 16.0000  sd = 0.0075
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.1)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0041
CI are -0.0240 and -0.0054
Tstat = -3.3438  df = 16.0000  sd = 0.0093
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.15)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.1151
CI are -0.0196 and 0.0024
Tstat = -1.6661  df = 16.0000  sd = 0.0110
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.2)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.3009
CI are -0.0184 and 0.0061
Tstat = -1.0690  df = 16.0000  sd = 0.0122
```

```
***************************************************
48 hour forecast (Cost-Loss Ratio = 0.25)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 1.0000
CI are -0.0139 and 0.0139
Tstat = 0.0000  df = 16.0000  sd = 0.0139
***************************************************
48 hour forecast (Cost-Loss Ratio = 0.3)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.4588
CI are -0.0187 and 0.0088
Tstat = -0.7592  df = 16.0000  sd = 0.0138
***************************************************
48 hour forecast (Cost-Loss Ratio = 0.35)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.4313
CI are -0.0199 and 0.0089
Tstat = -0.8073  df = 16.0000  sd = 0.0144
***************************************************
48 hour forecast (Cost-Loss Ratio = 0.4)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.3776
CI are -0.0214 and 0.0086
Tstat = -0.9075  df = 16.0000  sd = 0.0150
***************************************************
48 hour forecast (Cost-Loss Ratio = 0.45)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.2953
CI are -0.0235 and 0.0076
Tstat = -1.0820  df = 16.0000  sd = 0.0155
***************************************************
48 hour forecast (Cost-Loss Ratio = 0.5)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.2153
CI are -0.0259 and 0.0063
Tstat = -1.2904  df = 16.0000  sd = 0.0161
***************************************************
48 hour forecast (Cost-Loss Ratio = 0.55)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.1484
CI are -0.0283 and 0.0047
Tstat = -1.5185  df = 16.0000  sd = 0.0165
***************************************************
48 hour forecast (Cost-Loss Ratio = 0.6)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.0955
CI are -0.0310 and 0.0028
Tstat = -1.7715  df = 16.0000  sd = 0.0169
```

```
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.65)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.0610
CI are -0.0338 and 0.0009
Tstat = -2.0156  df = 16.0000  sd = 0.0173
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.7)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0347
CI are -0.0367 and -0.0016
Tstat = -2.3081  df = 16.0000  sd = 0.0176
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.75)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0208
CI are -0.0397 and -0.0038
Tstat = -2.5638  df = 16.0000  sd = 0.0180
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.8)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0119
CI are -0.0428 and -0.0062
Tstat = -2.8389  df = 16.0000  sd = 0.0183
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.85)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0071
CI are -0.0459 and -0.0085
Tstat = -3.0843  df = 16.0000  sd = 0.0187
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.9)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0041
CI are -0.0490 and -0.0110
Tstat = -3.3516  df = 16.0000  sd = 0.0190
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.95)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0025
CI are -0.0521 and -0.0134
Tstat = -3.5870  df = 16.0000  sd = 0.0194
```

CASE 3

```
November Hypothesis Testing Results for Global
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.05)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0182 and -0.0121
Tstat = -10.3500  df = 24.0000  sd = 0.0037
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.1)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0121 and -0.0054
Tstat = -5.4277  df = 24.0000  sd = 0.0041
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.15)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0099
CI are -0.0084 and -0.0013
Tstat = -2.7992  df = 24.0000  sd = 0.0044
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.2)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.1765
CI are -0.0062 and 0.0012
Tstat = -1.3927  df = 24.0000  sd = 0.0046
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.25)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 1.0000
CI are -0.0038 and 0.0038
Tstat = 0.0000  df = 24.0000  sd = 0.0047
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.3)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.6319
CI are -0.0048 and 0.0030
Tstat = -0.4852  df = 24.0000  sd = 0.0048
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.35)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.6209
CI are -0.0049 and 0.0030
Tstat = -0.5011  df = 24.0000  sd = 0.0049
```

```
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.4)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.4457
CI are -0.0056 and 0.0025
Tstat = -0.7753  df = 24.0000  sd = 0.0050
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.45)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.2348
CI are -0.0066 and 0.0017
Tstat = -1.2187  df = 24.0000  sd = 0.0051
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.5)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.0897
CI are -0.0078 and 0.0006
Tstat = -1.7685  df = 24.0000  sd = 0.0052
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.55)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0233
CI are -0.0093 and -0.0007
Tstat = -2.4222  df = 24.0000  sd = 0.0053
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.6)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0042
CI are -0.0111 and -0.0023
Tstat = -3.1646  df = 24.0000  sd = 0.0054
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.65)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0006
CI are -0.0130 and -0.0041
Tstat = -3.9473  df = 24.0000  sd = 0.0055
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.7)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0001
CI are -0.0150 and -0.0060
Tstat = -4.8081  df = 24.0000  sd = 0.0056
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.75)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0172 and -0.0080
Tstat = -5.6889  df = 24.0000  sd = 0.0056
```

```
**************************************************
24 hour forecast (Cost-Loss Ratio = 0.8)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0194 and -0.0102
Tstat = -6.6273  df = 24.0000  sd = 0.0057
**************************************************
24 hour forecast (Cost-Loss Ratio = 0.85)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0217 and -0.0124
Tstat = -7.5309  df = 24.0000  sd = 0.0058
**************************************************
24 hour forecast (Cost-Loss Ratio = 0.9)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0241 and -0.0146
Tstat = -8.4293  df = 24.0000  sd = 0.0059
**************************************************
24 hour forecast (Cost-Loss Ratio = 0.95)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0266 and -0.0170
Tstat = -9.2934  df = 24.0000  sd = 0.0060
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.05)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0216 and -0.0140
Tstat = -9.6655  df = 24.0000  sd = 0.0047
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.1)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0142 and -0.0062
Tstat = -5.2383  df = 24.0000  sd = 0.0050
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.15)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0052
CI are -0.0104 and -0.0020
Tstat = -3.0705  df = 24.0000  sd = 0.0052
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.2)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.0584
CI are -0.0086 and 0.0002
Tstat = -1.9876  df = 24.0000  sd = 0.0054
```

```
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.25)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 1.0000
CI are -0.0044 and 0.0044
Tstat = 0.0000  df = 24.0000  sd = 0.0054
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.3)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.1715
CI are -0.0078 and 0.0015
Tstat = -1.4094  df = 24.0000  sd = 0.0057
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.35)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.1244
CI are -0.0085 and 0.0011
Tstat = -1.5925  df = 24.0000  sd = 0.0059
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.4)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.0726
CI are -0.0094 and 0.0004
Tstat = -1.8778  df = 24.0000  sd = 0.0061
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.45)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0312
CI are -0.0105 and -0.0005
Tstat = -2.2890  df = 24.0000  sd = 0.0062
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.5)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0092
CI are -0.0120 and -0.0019
Tstat = -2.8330  df = 24.0000  sd = 0.0062
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.55)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0022
CI are -0.0136 and -0.0034
Tstat = -3.4304  df = 24.0000  sd = 0.0063
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.6)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0005
CI are -0.0153 and -0.0049
Tstat = -4.0199  df = 24.0000  sd = 0.0064
```

```
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.65)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0001
CI are -0.0172 and -0.0067
Tstat = -4.6823  df = 24.0000  sd = 0.0065
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.7)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0192 and -0.0085
Tstat = -5.3618  df = 24.0000  sd = 0.0066
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.75)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0213 and -0.0104
Tstat = -6.0302  df = 24.0000  sd = 0.0067
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.8)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0234 and -0.0124
Tstat = -6.6859  df = 24.0000  sd = 0.0068
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.85)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0255 and -0.0143
Tstat = -7.3371  df = 24.0000  sd = 0.0069
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.9)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0277 and -0.0163
Tstat = -7.9833  df = 24.0000  sd = 0.0070
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.95)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0300 and -0.0184
Tstat = -8.6211  df = 24.0000  sd = 0.0071
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.05)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0225 and -0.0147
Tstat = -9.8306  df = 24.0000  sd = 0.0048
```

```
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.1)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0144 and -0.0061
Tstat = -5.1269  df = 24.0000  sd = 0.0051
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.15)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0052
CI are -0.0110 and -0.0022
Tstat = -3.0727  df = 24.0000  sd = 0.0054
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.2)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0397
CI are -0.0095 and -0.0002
Tstat = -2.1747  df = 24.0000  sd = 0.0057
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.25)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 1.0000
CI are -0.0049 and 0.0049
Tstat = 0.0000  df = 24.0000  sd = 0.0060
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.3)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.0790
CI are -0.0089 and 0.0005
Tstat = -1.8344  df = 24.0000  sd = 0.0058
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.35)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0431
CI are -0.0096 and -0.0002
Tstat = -2.1363  df = 24.0000  sd = 0.0058
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.4)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0163
CI are -0.0108 and -0.0012
Tstat = -2.5840  df = 24.0000  sd = 0.0059
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.45)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0050
CI are -0.0123 and -0.0025
Tstat = -3.0939  df = 24.0000  sd = 0.0061
```

```
**************************************************
72 hour forecast (Cost-Loss Ratio = 0.5)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0013
CI are -0.0138 and -0.0038
Tstat = -3.6486  df = 24.0000  sd = 0.0062
**************************************************
72 hour forecast (Cost-Loss Ratio = 0.55)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0003
CI are -0.0155 and -0.0054
Tstat = -4.2933  df = 24.0000  sd = 0.0062
**************************************************
72 hour forecast (Cost-Loss Ratio = 0.6)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0174 and -0.0072
Tstat = -4.9578  df = 24.0000  sd = 0.0063
**************************************************
72 hour forecast (Cost-Loss Ratio = 0.65)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0193 and -0.0089
Tstat = -5.5892  df = 24.0000  sd = 0.0064
**************************************************
72 hour forecast (Cost-Loss Ratio = 0.7)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0212 and -0.0107
Tstat = -6.2500  df = 24.0000  sd = 0.0065
**************************************************
72 hour forecast (Cost-Loss Ratio = 0.75)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0232 and -0.0125
Tstat = -6.8923  df = 24.0000  sd = 0.0066
**************************************************
72 hour forecast (Cost-Loss Ratio = 0.8)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0253 and -0.0144
Tstat = -7.5434  df = 24.0000  sd = 0.0067
**************************************************
72 hour forecast (Cost-Loss Ratio = 0.85)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0273 and -0.0163
Tstat = -8.1653  df = 24.0000  sd = 0.0068
```

```
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.9)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0293 and -0.0182
Tstat = -8.7599  df = 24.0000  sd = 0.0069
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.95)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0314 and -0.0200
Tstat = -9.3368  df = 24.0000  sd = 0.0070
```

CASE 4

```
November Hypothesis Testing Results for 30Nto55N
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.05)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0305 and -0.0230
Tstat = -14.7270  df = 24.0000  sd = 0.0046
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.1)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0212 and -0.0126
Tstat = -8.0903  df = 24.0000  sd = 0.0053
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.15)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0001
CI are -0.0154 and -0.0058
Tstat = -4.5142  df = 24.0000  sd = 0.0060
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.2)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0203
CI are -0.0122 and -0.0011
Tstat = -2.4865  df = 24.0000  sd = 0.0068
****************************************************
24 hour forecast (Cost-Loss Ratio = 0.25)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 1.0000
CI are -0.0060 and 0.0060
Tstat = 0.0000  df = 24.0000  sd = 0.0074
```

```
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.3)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.2671
CI are -0.0098 and 0.0028
Tstat = -1.1361  df = 24.0000  sd = 0.0078
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.35)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.3422
CI are -0.0097 and 0.0035
Tstat = -0.9691  df = 24.0000  sd = 0.0082
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.4)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.3595
CI are -0.0101 and 0.0038
Tstat = -0.9341  df = 24.0000  sd = 0.0086
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.45)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.2569
CI are -0.0114 and 0.0032
Tstat = -1.1613  df = 24.0000  sd = 0.0090
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.5)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.1503
CI are -0.0131 and 0.0021
Tstat = -1.4859  df = 24.0000  sd = 0.0094
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.55)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.0736
CI are -0.0150 and 0.0007
Tstat = -1.8706  df = 24.0000  sd = 0.0097
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.6)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0286
CI are -0.0175 and -0.0011
Tstat = -2.3293  df = 24.0000  sd = 0.0102
***************************************************
24 hour forecast (Cost-Loss Ratio = 0.65)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0083
CI are -0.0204 and -0.0034
Tstat = -2.8793  df = 24.0000  sd = 0.0105
```

```
*****************************************************
24 hour forecast (Cost-Loss Ratio = 0.7)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0020
CI are -0.0237 and -0.0060
Tstat = -3.4615  df = 24.0000  sd = 0.0109
*****************************************************
24 hour forecast (Cost-Loss Ratio = 0.75)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0004
CI are -0.0273 and -0.0090
Tstat = -4.0868  df = 24.0000  sd = 0.0113
*****************************************************
24 hour forecast (Cost-Loss Ratio = 0.8)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0001
CI are -0.0310 and -0.0119
Tstat = -4.6522  df = 24.0000  sd = 0.0118
*****************************************************
24 hour forecast (Cost-Loss Ratio = 0.85)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0348 and -0.0151
Tstat = -5.2474  df = 24.0000  sd = 0.0121
*****************************************************
24 hour forecast (Cost-Loss Ratio = 0.9)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0390 and -0.0186
Tstat = -5.8465  df = 24.0000  sd = 0.0126
*****************************************************
24 hour forecast (Cost-Loss Ratio = 0.95)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0433 and -0.0223
Tstat = -6.4582  df = 24.0000  sd = 0.0129
*****************************************************
48 hour forecast (Cost-Loss Ratio = 0.05)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0366 and -0.0267
Tstat = -13.2238  df = 24.0000  sd = 0.0061
*****************************************************
48 hour forecast (Cost-Loss Ratio = 0.1)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0251 and -0.0147
Tstat = -7.9282  df = 24.0000  sd = 0.0064
```

```
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.15)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0001
CI are -0.0189 and -0.0076
Tstat = -4.8309  df = 24.0000  sd = 0.0070
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.2)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0038
CI are -0.0155 and -0.0034
Tstat = -3.1990  df = 24.0000  sd = 0.0075
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.25)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 1.0000
CI are -0.0064 and 0.0064
Tstat = 0.0000  df = 24.0000  sd = 0.0079
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.3)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.0976
CI are -0.0128 and 0.0011
Tstat = -1.7240  df = 24.0000  sd = 0.0086
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.35)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.1099
CI are -0.0130 and 0.0014
Tstat = -1.6601  df = 24.0000  sd = 0.0089
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.4)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 0.0819
CI are -0.0140 and 0.0009
Tstat = -1.8162  df = 24.0000  sd = 0.0092
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.45)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0390
CI are -0.0159 and -0.0004
Tstat = -2.1840  df = 24.0000  sd = 0.0095
**************************************************
48 hour forecast (Cost-Loss Ratio = 0.5)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0118
CI are -0.0182 and -0.0025
Tstat = -2.7253  df = 24.0000  sd = 0.0097
```

```
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.55)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0028
CI are -0.0210 and -0.0049
Tstat = -3.3320  df = 24.0000  sd = 0.0099
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.6)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0006
CI are -0.0239 and -0.0074
Tstat = -3.9303  df = 24.0000  sd = 0.0101
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.65)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0001
CI are -0.0269 and -0.0103
Tstat = -4.6021  df = 24.0000  sd = 0.0103
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.7)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0304 and -0.0134
Tstat = -5.3048  df = 24.0000  sd = 0.0105
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.75)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0339 and -0.0165
Tstat = -5.9635  df = 24.0000  sd = 0.0108
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.8)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0375 and -0.0197
Tstat = -6.6438  df = 24.0000  sd = 0.0110
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.85)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0413 and -0.0231
Tstat = -7.3080  df = 24.0000  sd = 0.0112
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.9)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0452 and -0.0266
Tstat = -7.9659  df = 24.0000  sd = 0.0115
```

```
****************************************************
48 hour forecast (Cost-Loss Ratio = 0.95)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0493 and -0.0303
Tstat = -8.6474  df = 24.0000  sd = 0.0117
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.05)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0410 and -0.0317
Tstat = -16.2411  df = 24.0000  sd = 0.0057
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.1)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0271 and -0.0168
Tstat = -8.8459  df = 24.0000  sd = 0.0063
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.15)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0206 and -0.0095
Tstat = -5.5929  df = 24.0000  sd = 0.0069
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.2)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0007
CI are -0.0170 and -0.0052
Tstat = -3.8602  df = 24.0000  sd = 0.0073
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.25)
H = 0 (Cannot Reject Null Hypothesis)
The sample means may not be significantly different (.05 Level).
Significance (P-value) = 1.0000
CI are -0.0067 and 0.0067
Tstat = 0.0000  df = 24.0000  sd = 0.0083
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.3)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0290
CI are -0.0146 and -0.0009
Tstat = -2.3231  df = 24.0000  sd = 0.0085
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.35)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0315
CI are -0.0152 and -0.0008
Tstat = -2.2835  df = 24.0000  sd = 0.0089
```

```
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.4)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0214
CI are -0.0167 and -0.0015
Tstat = -2.4620  df = 24.0000  sd = 0.0094
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.45)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0086
CI are -0.0188 and -0.0030
Tstat = -2.8640  df = 24.0000  sd = 0.0097
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.5)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0029
CI are -0.0213 and -0.0050
Tstat = -3.3196  df = 24.0000  sd = 0.0101
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.55)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0005
CI are -0.0244 and -0.0078
Tstat = -4.0141  df = 24.0000  sd = 0.0102
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.6)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0001
CI are -0.0278 and -0.0109
Tstat = -4.7484  df = 24.0000  sd = 0.0104
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.65)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0312 and -0.0139
Tstat = -5.3965  df = 24.0000  sd = 0.0106
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.7)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0347 and -0.0171
Tstat = -6.0549  df = 24.0000  sd = 0.0109
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.75)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0385 and -0.0203
Tstat = -6.6949  df = 24.0000  sd = 0.0112
```

```
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.8)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0422 and -0.0236
Tstat = -7.2886  df = 24.0000  sd = 0.0115
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.85)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0461 and -0.0269
Tstat = -7.8640  df = 24.0000  sd = 0.0118
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.9)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0500 and -0.0303
Tstat = -8.4006  df = 24.0000  sd = 0.0122
****************************************************
72 hour forecast (Cost-Loss Ratio = 0.95)
H = 1 (Reject Null Hypothesis)
The sample means are significantly different (.05 Level).
Significance (P-value) = 0.0000
CI are -0.0539 and -0.0336
Tstat = -8.8863  df = 24.0000  sd = 0.0125
```

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D: EXAMPLE ORM WORKSHEETS

Appendix D consists of figures of pages (Figures 37 and 38) of an ORM worksheet from the 101 ARW from the Maine Air National Guard (Lt. Col. Andrew Marshall, 2005, personal communication).

101 ARW Operations Risk Assessment Matrix

Date:__________ **Sortie Number:**______ **SOF Name:**________________ **AC Name:**________________

	Low	Pts	Medium	Pts	High	Pts
Show Time	0600 - 1630L	0	0430 - 0600L or 1630 - 1900L	5	1900 - 0430L	15
Crew Duty Time (Any crewmember)	Less than 8 hours	0	8 to 12 hours	5	More than 12 hours	15
Fatigue	Well rested	0	One crewmember tired	5	Two or more tired	15
Stress	Normal Stress	0	One crewmember stressed	5	Two or more stressed	15
Last Sortie (Any crewmember)	Less than 2 weeks	0	More than 2 weeks but less than 4 weeks	5	4 weeks or more	10
Experience (In Primary Position)	ALL more than 200 hrs in -135	0	ONE under 200 hrs or ANY non-current	5	Upgrade/diff training or more than one under 200 hours	15
Crew	Basic Crew	0	Augmented 3P (2 ACs & 2 BOs)	-10	Augmented 4P (2 ACs, 2 Navs, 2 BOs)	-15
Priority (See reverse)	Peacetime Training (3B, 4A, 5A)	0	Peacetime HHD or ORI / Exercise (2A, 2B, 2C, 3A)	10	POTUS, NAOC, Special Ops, Wartime (1A, 1B)	25
Complexity	Cargo or Pax, ERCC	5	Deployment or XC, Hazardous Cargo	10	DV Pax, TTF Mission, Aeromedical	20
Tactics	Two-ship VMC, Local VFR Training	5	Two-ship IMC, Multi-ship VMC, Multi-Fighter Anchor,	10	Multi-ship IMC, Multi-service Ops, OCONUS Theater Ops, Tactical/Chemical Threats	15
Delay / Changes	None	0	Less than two hours, Minor re-planning	5	Two hours or more, Major re-planning	15
Airfields	KBGR	0	Off-station familiar	5	Unfamiliar field	10
Airfield Conditions (Worst Case)	Wet runway, Night, Crosswind or gusts > 10, Precip, Cold WX Proc.	5	Ceil/Vis below 500' or 1, > 30 min night transition, RCR 7 to 9, Crosswind or gusts > 15, Birdwatch Moderate	10	Ceil/Vis below 500' or 1 in precip, RCR 6 or less, Crosswind or gusts > 20, Radar / ATC issues	20
Route of Flight	CONUS Familiar	0	CONUS Unfamiliar, OCONUS Familiar	5	OCONUS Unfamiliar	15
Enroute Conditions (Worst Case)	None	0	Light icing or turbulence, Isolated or Few thunderstorms	5	Mod icing or turbulence, Scattered or Numerous Thunderstorms, Radar / ATC issues	15
Aircrew Arming	Not required: Altn not reqd, No Stops, No Pax, BGR Transition	0	Required: (OG or SOF can waive) Alternate required, Planned stop, Unit Pax, Off-station transition	5	Required: Transition field or Alternate under FPCON C or D, Non-unit Pax	10

TOTAL POINTS:	RISK LEVEL:	ACTIONS:
	0 to 49 = Low Risk	Crew Reviews Risks
	50 to 80 = Caution	Crew and SOF Review Risks
	81 or More = High Risk	Crew and SOF Review Risks - Attempt to Reduce Risk Level

I certify that the aircrew is current and qualified, have read and signed the FCIF, have received applicable Intel/Threat briefings. I have reviewed this worksheet and discussed it with the crew if required.

____________________________, Supervisor of Flying

ORM Worksheet.doc Nov 2003 OPR: OGV

Figure 37. ORM Worksheet Page 1.

AIR REFUELING SUPPORT PRIORITIES

Extracted from AFI 11-221 (Air Refueling Management. 1 Nov 1995) Attachment 1

25 POINTS – HIGH*

Priority 1A.
- 1A1 - Presidential-directed and operational National Airborne Ops Center (NAOC)
- 1A2 - Wartime or JCS-designated contingency combat support.
- 1A3 - Special ops support and other programs approved by the President for top national priority.

Priority 1B.
- 1B1 - Contingency deployments and SECDEF/JCS-directed special missions.
- 1B2 - Counterdrug and operational reconnaissance.

NOTE: Priority 1 missions are eligible for tanker spare aircraft or 24-hour slip capability, when available.

10 POINTS – MEDIUM*

Priority 2A.
- 2A1 - Nonscheduled JCS-directed operational deployments.
- 2A2 - JCS-directed exercises requiring air refueling to meet JCS objectives.
- 2A3 - Over water deployments or deployment of aircraft tasked for Priority 1 missions.

Priority 2B.
- 2B1 - Foreign Military Sales (FMS) support.
- 2B2 - Aircraft test ops.
- 2B3 - Over water redeployments. Redeployment of Priority 1-tasked aircraft. Scheduled aircraft swap out deployments.

Priority 2C.
- 2C1 - JCS-directed exercises requiring air refueling to meet MAJCOM, NAF, or wing objectives.
- 2C2 - Employment missions in MAJCOM-directed exercises/ops. MAJCOM/NAF/Wing-directed over water deployments.
- 2C3 - Predeployment qual training.

Priority 3A.
- 3A1 - MAJCOM/NAF/Wing-directed redeployments or NAF-directed exercises/ORIs.
- 3A2 - Intratheater deployments and redeployments.

0 POINTS - LOW

Priority 3B.
- 3B1 - CCTS/RTU, requal/upgrade training, when air refueling training is accomplished during the mission.
- 3B2 - Wing-directed exercises and evaluations.

Priority 4A.
- 4A1 - Missions launched to satisfy USAF, USN, and other DoD agency training requirements.

Priority 5A.
- 5A1 - Unit to unit scheduled non-allocated air refueling.

*Medium and High priority missions on this list are Operational Missions

Figure 38. ORM Worksheet Page 2.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Air Force Doctrine Document 1 (AFDD 1), Nov 17, 2003: *Air Force Basic Doctrine*.

Air Force Instruction 15-128 (AFI 15-128), cited 2005: Air and Space Weather Operations Roles and Responsibilities. [Available online at: http://www.e-publishing.af.mil/pubfiles/af/15/afi15-128/afi15-128.pdf], (Accessed Nov 17, 2005).

Air Force Manual 15-129 (AFMAN 15-29), cited 2005: Air and Space Weather Operations – Processes and Procedures. [Available online at http://www.e-publishing.af.mil/pubfiles/af/15/afman15-129/afman15-129.pdf], (Accessed Nov 17, 2005).

Air Force Pamphlet 90-902 (AFPAM 90-902), cited 2006: Operational Risk Management (ORM) – Guidelines and Tools. [Available online at http://www.e-publishing.af.mil/pubfiles/af/90/afpam90-902/afpam90-902.pdf], (Accessed Feb 22, 2006).

Air Force Weather Agency Technical Note 98-002 (AFWA TN98-2002), cited 2005: Meteorological Techniques. [Available online at: https://wwwmil.offutt.af.mil/afwadnt/doctrace.asp?file=TN_98002_Complete.zip&tag=TN 98-002], (Accessed Oct 15, 2005).

AMS 2003: Support for automated observations from U.S. commercial aircraft. *Bulletin of the American Meteorological Society*, **84**, 515-517.

Anderson, T. W., J. D. Finn, 1996: *The New Statistical Analysis of Data*. Springer-Verlag, 712 pp.

Anthes, R. A., 1986: The general question of predictability. *Mesoscale Meteorology and Forecasting*, American Meteorological Society, Boston, MA, 636–656.

Barksdale Operational Weather Squadron (OWS), cited 2005: OWS Produced Human Turbulence Chart. [Available online at: https://ows.barksdale.af.mil], (Accessed Oct 15, 2005).

Brooks, H., Doswell III, C., 1993: New technology and numerical weather prediction - a wasted opportunity? *Weather*, **48**, 173-177.

Cornman, L. B., C. S. Morse, and G. Cunning, 1995 Real-time estimation of atmospheric turbulence severity from in-situ aircraft measurements. *Journal of Aircraft*, **1**, 171-177.

Department of Defense Form 175-1 (DD175-1), cited 2005: Flight Weather Briefing. [Available online at: http://www.dtic.mil/whs/directives/infomgt/forms/forminfo/forminfopage38.html], (Accessed Nov 17, 2005).

Dutton, M.J.O., 1980: Probability forecasts of clear-air turbulence based on numerical model output. *Meterol. Mag.*, **109**, 293-310.

Eckel, F. A.,1998: Calibrated probabilistic quantitative precipitation forecasts based on the MRF ensemble. M.S. thesis, Graduate School of Engineering, Air Force Institute of Technology, Air University, Air Education and Training Command. 133 pp. [Available from the Defense Technical Information Center]

_________, 2003: Effective Mesoscale, Short-Range Ensemble Forecasting. Ph.D. dissertation, University of Washington, Seattle, WA, 224 pp. [Available from University of Washington, Seattle, WA 98195].

_________, and C. F. Mass, 2005: Aspects of effective mesoscale, short-range ensemble forecasting**.** *Weather and Forecasting,* **20**, 328–350.

Ellrod, G. P. and D. I. Knapp, 1992: An objective clear-air turbulence forecasting technique: verification and operational use. *Weather and Forecasting*, **7**, 150–165.

ESRL/GSD, cited 2005a: FSL Aircraft Data (ACARS/AMDAR) Information. [Available online at http://amdar.noaa.gov/FAQ.html.], (Accessed Oct 15, 2005).

_________, cited 2005b: Ancillary information for Automated Meteorological Reports from Commercial Aircraft. [Available online at http://acweb.fsl.noaa.gov/bams/ancillary/.], (Accessed Oct 15, 2005).

Frehlich R. and R. Sharman. 2004: Estimates of Turbulence from Numerical Weather Prediction Model Output with Applications to Turbulence Diagnosis and Data Assimilation. *Monthly Weather Review*, **132**, 2308–2324.

Hamill T. M. and S. J. Colucci. 1997: Verification of Eta–RSM Short-Range Ensemble Forecasts**.** *Monthly Weather Review,* **125**, 1312–1327.

Joint Air Force and Army Information Network (JAAWIN), cited 2005: AFWA Produced Model-derived Turbulence Chart. [Available online at https://weather.afwa.af.mil], (Accessed Nov 22, 2005).

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341 pp.

Lewis, J. M., 2005: Roots of Ensemble Forecasting. *Monthly Weather Review,* **133**, 1865-1885.

Lorenz, E. N., 1963: Deterministic Nonperiodic Flow. *Journal of Atmospheric Sciences*, **20**, 130-141.

MacCready Jr., P. B., 1964: Standardization of gustiness values from aircraft. *Journal of Applied Meteorology*, **3**, 439–449.

Moninger, W. R., R. Mamrosh, and P. Pauley, 2003: Automated meteorological reports from commercial aircraft. *Bulletin of the American Meteorological Society*, **84,** 203-216.

Murphy, A. H., R. W. Katz, R. L. Winkler, and W. Hsu., 1985: Repetitive Decision Making and the Value of Forecasts in the Cost-Loss Ratio Situation: A Dynamic Model. *Monthly Weather Review*, **113**, 801–813.

Palmer, T. N., 2002: The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quarterly Journal of the Royal Meteorological Society*, **128**, 747-774.

Schwartz, B., 1996: The Quantitative Use of PIREPs in Developing Aviation Weather Guidance Products. *Weather and Forecasting*, **11**, 372–384.

Tebaldi, C. D. Nychka, B. G. Brown, and R. Sharman, 2002: Flexible discriminant techniques for forecasting clear-air turbulence. *Environmetrics*, **13**, 859-878doi:10.1002/env.562.

Toth, Z., E. Kalnay, S. Tracton, R. Wobus, and J. Irwin, 1997: A Synoptic Evaluation of the NCEP Ensemble. *Weather and Forecasting*, **12,** 152-153.

_________, Vannitsem, S., 2005: Model Errors and Ensemble Forecasting. *NCEP Website*, [Available online at: http://wwwt.emc.ncep.noaa.gov/gmb/ens/papers/ecmwf-finalwordweb.htm], (Accessed Jun 3, 2005).

UCAR, cited 2005a: RAP annual report 2003. U*CAR Website*. [Available online at: http://www.rap.ucar.edu/asr2003/atmospheric-turbulence.html], (Accessed Aug 15, 2005)

_________, cited 2005b: RAP annual report 2004. U*CAR Website*. [Available online at: http://www.asr.ucar.edu/2004/RAP/atmospheric-turbulence.htm#2], (Accessed Aug 15, 2005).

_________, cited 2005c: Ensemble Forecasting Explained. *COMET MetEd Website.* [Available online at: http://meted.ucar.edu/topics_nwp.php], (Accessed Aug 15, 2005).

Weiss, S. J., 1977: Objective verification of the severe weather outlook at the National Severe Storms Forecast Center. Preprints, 10th *Conf. on Severe Local Storms*, Omaha, Nebraska, American Meteorological Soc., 395-402.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2d ed. Academic Press, 627 pp.

Wilson, L. J. W. R. Burrows, and A. Lanzinger, 1999: A Strategy for Verification of Weather Element Forecasts from an Ensemble Prediction System. *Monthly Weather Review*, **127**, 956–970.

Zhou, B., J. Du, J. McQueen, G. Dimego, G. Manikin, B. Ferrier, Z. Toth, H. Juang, M. Hart, and J. Hann, 2004: An Introduction to NCEP SREF Aviation Project, *11th Conference on Aviation, Range, and Aerospace*, Hyannis, Amer. Meteor. Soc.

Zhu, Y., Z. Toth, R. Wobus, D. Richardson, and K. Myline, 2002: The Economic Value of Ensemble-Based Weather Forecasts. Bulletin *of the American Meteorological Society*. Vol. 83, No. 1, pp. 73–83.

BIBLIOGRAPHY

Bougeault, P., 2003: The WGNE survey of verification methods for numerical prediction of weather elements and severe weather events. *WMO Report on Research Activities in Atmosphere and Oceanic Modeling,* Report Number 33, Appendix.

Colson, D. and H. A. Panofsky, 1965: An index of clear-air turbulence. *Quart. J.Roy. Meteor. Soc*., **91**, 507-513.

Fleming, R. J., 1996: The Use of Commercial Aircraft as Platforms for Environmental Measurements. *Bulletin of the American Meteorological Society*. **77**, No. 2229–2242.

Kao, S.-K. and A. H. Sizoo, 1966: Analysis of clear air turbulence near the jet stream. *Journal of Geophysical Research*. **71**, 3799-3805.

Kronebach G. W., 1964: An automated procedure for forecasting clear-air turbulence. *Journal of Applied Meteorology,* **3** 119-125.

Mullen, S. L. and D. P. Baumhefner, 1989: The impact of initial condition uncertainty on numerical simulations of large-scale explosive cyclogenesis. *Monthly Weather Review,* **117**, 2800-2821.

Raiffa, H., 1970: *Decision Analysis: Introductory Lectures on Choices under Uncertainty*. Addison-Wesley Publishing Company, 309 pp.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Air Force Weather Technical Library
   Air Force Combat Climatology Center
   Asheville, North Carolina

4. Dr. Chuck Wash
   Naval Postgraduate School
   Monterey, California

5. Dr. Patrick Harr
   Naval Postgraduate School
   Monterey, California

6. Dr. Philip Durkee
   Naval Postgraduate School
   Monterey, California

7. Dr. Russell Elsberry
   Naval Postgraduate School
   Monterey, California

8. Mr. Robert Creasey
   Naval Postgraduate School
   Monterey, California

9. Major Tony Eckel, USAF
   Air Force Weather Agency
   Offutt AFB, Nebraska

10. Dr. Eva Regnier
    Defense Resources Management Institute
    Monterey, California

11. Dr. Craig Bishop
    Naval Research Lab
    Monterey, California